Energy Research & Social Science 52 (2019) 221–235

Contents lists available at ScienceDirect

Energy Research & Social Science

journal homepage: www.elsevier.com/locate/erss

What makes a cookstove usable? Trials of a usability testing protocol in Uganda, Guatemala, and the United States

Nicholas D. Moses, Nordica A. MacCarty*

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

ARTICLE INFO

Keywords:
Evaluation
Biomass
Ethnographic

ABSTRACT

Improved cookstoves, which are designed to address the health and environmental damage caused by inefficient biomass cooking and heating, are generally engineered to prioritize technical emissions and efficiency improvements over practical user cooking needs. As a result, improved stoves are often only used alongside traditional technologies, or are rejected by cooks entirely, and may fall short in efforts to address health and environmental problems even if they are technically capable of doing so. A testing protocol for cookstove usability was developed by the authors to help stove designers and implementers evaluate user needs for a given stove technology and cooking culture, and balance them with technical design criteria more effectively. The protocol is based on established usability practices and includes ethnographic testing methods to increase validity in cross-cultural testing applications. This paper discusses evaluative trials of the protocol in 10 rural and urban households and 2 institutional kitchens near Lira, Uganda, 4 rural households in Antigua, Guatemala, and 11 stove designs in a US laboratory, as well as preliminary results from a usability study on 20 rural households near Gulu, Uganda by an international NGO. These trials, along with feedback from test administrators in Uganda and Guatemala, demonstrated that the protocol is a viable tool for increasing the understanding of cookstove usability, though more work is needed to improve its validity and connect the results to effective stove design and selection decisions, as well as health and environmental impact. As a result of this research, various improvements were made to the protocol, and opportunities for further validation, expansion, and improvement were identified. In addition, methods used by the protocol elicited related data from participants regarding their attitudes towards improved cookstoves and the relative importance of reducing air pollution and fuel use in the larger context of their lives. Finally, this research highlights the value of an interdisciplinary approach, in this case based in anthropology and engineering, to increase the accessibility of user input in international development more broadly.

1. Introduction

Household air pollution from sources including inefficient biomass cooking technologies has been estimated to kill 4 million people annually [1]. Unsustainable wood fuel harvesting practices have also been shown to cause environmental damage in some regions [2], while emissions from traditional fires and stoves also contribute to local and global climate change [3]. Up to one billion dollars is spent each year addressing these issues through improved cookstove programs [4], which in theory have the ability to solve the health and environmental issues associated with burning biomass [5]. Some studies have found minimal impact from the tens of millions of improved stoves that have been produced, however, often because stoves are not used, or are used alongside inefficient, polluting traditional methods [5,6]. One significant reasons for this shortcoming is low usability; if improved stove designs do not fully meet the user's cooking needs, users will supplement them with more practical and usable traditional stoves, or will not bother obtaining them at all [5–9].

Usability testing, or the study of how efficiently and effectively a product can be used in a certain context, has become an important and well-studied part of design fields across many sectors in high-income countries. These include healthcare systems, web and software, and consumer product design [10]. Research into usability and user needs is especially helpful when designers have little previous understanding of a user or context and cannot draw on their own experience to develop appropriate designs, such as in international development where users and designers often come from different countries and cultures [11]. Despite the common application of usability testing in design for high-

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

https://doi.org/10.1016/j.erss.2019.02.002
Received 20 June 2018; Received in revised form 4 January 2019; Accepted 5 February 2019

income countries, as well as its value for design in an international development, there is little published research on practices for defining and measuring usability in design for low-income nations. In the case of improved cookstoves and more broadly, this may be due in part to the inherent conflict between usability and technical performance goals. Technical criteria are more familiar to designers and engineers, easier to measure, and more likely to be promoted by internationally recognized tests and standards [7,12], and therefore may be more likely to receive attention. While cookstove usability and user needs have been considered by several researchers, and there has been a call to certify cookstoves for their given context of use, instead of through laboratory tests that isolate them from real cooks and cooking cultures [13], there has historically been no standardized or broadly applicable testing protocol available to help practitioners evaluate user design criteria.

A cookstove usability testing protocol [14] was developed by the authors to address this gap by offering a practical starting point for the evaluation of cookstove usability. The protocol was first trialed in and around Lira, in Northern Uganda, in the summer of 2017 with the assistance of International Lifeline Fund (ILF), a US-based international NGO operating programs including cookstove design and dissemination in that region. A second trial was conducted in the summer of 2018 as a part of an Oregon State University (OSU) Humanitarian Engineering field course in 4 households in Antigua, Guatemala. The purpose of these field studies was to provide preliminary validation and identify opportunities for improvement through the reflective process of administering the protocol, as well as through input from the Ugandan ILF staff and OSU students asked to administer the tests. The trials were not meant to thoroughly evaluate the usability of a particular model of stove.

A mix of urban and rural households, as well as two institutional kitchens, were included in these trials to assess the protocol with a range of different cooking technologies and contexts. Testing took place in kitchens during the preparation of the main daily meal, and cooks were asked not to modify their meals or cooking practices. Human subject research in Lira and Antigua were conducted with oversight from the OSU Institutional Review Board under study number 7257.

A preliminary assessment of an independent usability study conducted by Aid Africa, a US-based NGO, on 20 of their constituents near Gulu, Uganda, is also discussed in this paper. This work was done to provide an evaluation of Aid Africa's existing stove and stove programs, and the data and feedback were shared with the authors for their input and interpretation of the usability results, as well as to provide another opportunity to validate and improve the protocol.

Finally, the portion of the protocol containing quantitative and objective stove measurements was trialed in the laboratory of the Aprovecho Research Center in the United States on 11 common cookstove models. This was done to further improve this portion of the protocol, validate it on a wider variety of stove designs, and to compare the likely value of results collected from laboratory versus field testing. As a disclosure, the authors have a working relationship with the Aprovecho research center; the corresponding author being a past employee.

2. Background

2.1. Assessment of usability protocols

The assessment of a usability protocol is critical to determine its effectiveness and identify areas for improvement before effort is invested in its use. The tools and methods available for the assessment of a new protocol have increased over the last few decades with the emergence of usability testing as a distinct area practice, though the lack of standardization and the variety of approaches used to measure usability has made the effective judgment and comparison of protocols challenging [15,16]. Protocols may be assessed comparatively, where each is administered for a given application, then duplicate and unique usability problems identified are analyzed to determine which is able to elicit the most [16]. Alternatively, objective, quantitative standards by which to judge protocols have been devised and shown to be effective for consumer products [17]. The value in selecting a single most effective protocol for an application has been challenged, however, as has the validity of an assessment method focused on the quantity of defined usability problems identified, as opposed to their impact on design or user experience [16,18]. Similarly, standards-based assessments are limited to the applications and contexts they were designed for.

In the absence of comparable usability protocols or relevant protocol standards, there is no established method to assess a new protocol, though additional guidance has been provided by several researchers. Hartson and Castillo suggest examining usability problems in a real-world context with real users as a way to enhance the assessment of a protocol [19], and Hornbæk advocates for asking designers whether and how the results of usability testing improve their design work, which is often the ultimate goal of usability testing [16]. In other words, they suggest including both the users of a product and the users of the eventual usability testing results – in this case cooks and cookstove designers or implementers, respectively – in the validation and refinement of a protocol. In addition, the qualitative, expert-based assessment of the data produced by a protocol has been suggested as necessary for effective characterization of a protocol, in part because any quantitative data produced during a preliminary, evaluative study are unlikely to be sufficient for statistically significant results [15].

2.2. Cross-cultural and field testing methods and considerations

Multiple usability researchers and practitioners have incorporated ethnographic methods into usability testing protocols to add context and depth to their data [20–22]. According to Kiewe, while ethnographic methods do not necessarily follow a single protocol and may be adapted to specific research needs, in this context they typically include elements of interviews, observation of people and activities, and documentation of physical artifacts [22]. The necessity of compromise between traditional anthropological ethnography, which often requires a year or longer and may begin without a rigid research plan, and the practical realities and constraints of non-academic applications ethnographic methods has also been studied. "Rapid ethnography," or a limited use of ethnographic and other methods spanning just hours, days, or weeks has become an increasingly popular approach in industries such as human-computer interaction, which often incorporates usability research [23], and public health and international development, where approaches such as Scrimshaw and Hurtado's "Rapid Assessment Procedures" [24] and Bentley's "Rapid Ethnographic Assessment" [25] were described in the 1980's.

Compared to traditional ethnography, Millen states that this work typically includes features such as a limited research scope, the use of key informants and multiple observers, and interactive observation methods [23]. Pink and Morgan add mixed-method data collection, multi-disciplinary teams, and an emphasis on applying findings to interventions as features of rapid ethnographic methods [26]. According to these researchers and others, rapid ethnographic methods are not a substitute nor a replacement for traditional ethnography, but an efficient application of ethnographic methods when time and resources are limited, yet a foundational understanding of users' culture and context is required [23,26–28]. They are especially applicable when researchers are unfamiliar with the people they are studying [24]. Similarly, Isaacs explores the challenges of convincing managers and others who are often unfamiliar with ethnographic methods of the value of investing time in these methods early in a design process. She concludes that rapid ethnography offers significant benefits to product and system designers, yet avoids requiring so much investment in unfamiliar techniques that they are deterred from applying ethnographic methods entirely [29]. The evaluation of the proposed cookstove usability protocol differed from conventional protocol evaluations in that it was conducted primarily in the field, and in cross-cultural settings, where the researchers and participants were from different cultures and in most cases did not speak the same language. This work therefore lent itself to a "rapid" ethnographic approach to better bridge cultural and language differences, accommodate work in uncontrolled field environments [30] compared to traditional protocol validation methods, and to be conducted within a limited time frame compared to traditional ethnography.

Another key element of cross-cultural test administration is the identity of the test administrators. Herman notes that unobtrusive observation and testing will reduce unwanted effects from the test administrator's presence in cross-cultural usability testing [31], and Hornbæk adds that a well-qualified test administrator who is a subject matter expert will greatly enhance the quality of usability testing results [16].

2.3. Qualitative approaches and other relevant work in cookstove and energy transition studies

Multiple researchers have identified and demonstrated untapped value in the use of ethnographic and other qualitative methods in cookstove and energy transition studies, which are often designed from an engineering or epidemiological perspective and tend to focus on quantitative measurements. Loo et. al demonstrated the effective evaluation of cooks' perceptions of factors related to usability including ease of use, functionality, and perceived health impacts with a qualitative research approach [32]. According to Jagadish and Dwivedi, ethnographic methods also have the potential to characterize cooking preferences at the household level, including usability criteria such as cleanliness, perceived smoke emissions, taste, and seasonal factors [33], while Iessa et al. emphasize that ethnographic methods allow for issues such as cooking and cookstove adoption to be understood from a local, instead of external, perspective [34]. Chatti et al. add that ethnographic methods are in fact necessary to understand local practices and perspectives [35], which is suggested by Hartson & Castillo for usability protocol evaluation [19].

More broadly, qualitative and ethnographic approaches are able to elicit relevant details such as the rationale behind stove stacking and other causes of program failures; information that also is of critical value to a researcher or practitioner testing cookstove usability, and not necessarily addressed by standard protocols or practices [35]. Herington et al. use the lens of social practice to understand the various actors involved in cooking energy transitions in India and make recommendations for programs and policies [36], while Davies has used ethnography to characterize the larger systems in which cookstove and cookstove organizations exist, and how strategies such as warranties and product standards have complex and sometimes adverse effects on local people and economies [37].

Engineers working in international development have begun to adopt qualitative and ethnographic methods, as well. Wood and Mattson note that an ethnographic approach offers designers a tool to navigate unfamiliarity with culture and language, as well as to overcome assumptions and more accurately understand a user's needs and context [11]. Similarly, Adkins et al. emphasize the value of combining quantitative tests with qualitative surveys to better understand usability and user preference [38].

Several researchers have emphasized the complexity of energy transitions and technology adoption decisions, as well. Jurisoo, Lambe, and Osborne highlight the importance of diverse factors such as tax incentives, the availability of fuel, etc. [39], Otte and Malakar, Greig, and van de Fliert add the consideration of cultural factors such as gender norms and traditional income generation strategies [40,41], and Akinitan, Jewitt, and Clifford argue that traditional norms and taboos can be powerful influences in technology adoption decisions [42]. Malakar adds that individual decisions regarding fuel preferences cannot be well-understood out of the larger context of all of the factors that influence their decision making [43], while Wang and Bailis suggest that in some circumstances improved cookstoves can actually alleviate perceived marginalization caused by traditional stoves among the lower classes [44]. Others have modelled and quantified energy transition phenomena, as well. Miremadi, Saboohi, and Jacobsson created a holistic set of indicators by which to quantify energy innovation [45], and Kar and Zerrifi developed an interdisciplinary framework to test energy transition processes [46]. All of these articles highlight the inherent complexity of and diverse, sometimes competing factors that influence energy transitions, which underscores the value of a holistic approach in understanding these phenomena.

3. Cookstove usability protocol

The cookstove usability testing protocol developed by the authors is described here to give the reader enough information to understand what it entails and how it was employed in the field and laboratory studies conducted. In the interest of space the full protocol, which includes 15 pages of instruction and explanation of the scoring for each test and an 11 page template for recording test data, is not included here. An article dedicated to the methodology underlying the test design, scoring, and the selection of criteria has been accepted for publication in *Energy for Sustainable Development* [47], and the full protocol is available online from Oregon State University [14].

This usability protocol is meant to allow designers and implementers to evaluate stove features and cooking performance metrics that relate to usability, as well as a cook's perception of the usability of their stove. This information can allow practitioners to compare an "improved" stove to a traditional stove to determine whether or not it is likely to be appropriate and successful in a given context compared to a baseline stove, to select the best option from various stove models for a given project, or to identify opportunities for improvement in a stove or stove program design. The protocol consists of a set of detailed instructions for use and an explanation of scoring, a template for collecting field data, and a Microsoft Excel spreadsheet to store and analyze test results.

Usability criteria were identified from a review of existing cookstove literature related to user preferences and usability [6,48–52] and organized into six main categories (described in Table 1), each with several sub-criteria, to provide both high-level and more detailed assessments of usability in terms of the "effectiveness, efficiency and satisfaction" with which a stove meets user needs in a given context [53]. While most criteria were repeated in multiple literature sources and are likely applicable in a range of cooking cultures, the protocol also borrows from expanded definitions of usability to include elements of user needs solicitation and space for test administrators to add criteria, since unlike in traditional usability testing, the needs of a cook may not be fully defined by a test administrator before testing begins. It should be noted that the protocol only includes user perceptions of safety and durability, as opposed to objective measurements of potential hazards or longevity, to avoid duplicating the existing cookstove safety and durability protocols created by Johnson and Bryden [54] and Colorado State University [55], respectively.

The test methods used to evaluate these criteria are based in established usability testing practices from industries such as healthcare, computer user interface, and consumer product design, as well as relevant ethnographic methods from usability testing and related application. These methods collectively include physical measurements of cookstove and cooking parameters, a Likert-scale survey and interview of the cook, and direct observation of a routine cooking event, which follows Kiewe's recommendations for integrating ethnographic methods into usability testing [22]. Survey questions are typically paired to elicit both a cook's perception of a given sub-criterion, as well as the relative importance of that sub-criterion so that a weighted average score may be calculated for each of the six main usability

Table 1
Cookstove usability criteria and sub-criteria.
Source: Moses and MacCarty [14]

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				Insect repellant
	Cooking speed	Tending/refueling effort	Long-term maintenance	Perceived smoke exposure	Lighting
	Versatility				Portability
		Fuel feed entry size		Perceived durability	Food drying/ smoking
		Visibility of fire		Pot soot deposits	Water heating
		Ease of lighting		Cooking height	*(Additional needs may be added by the test administrator)*
		Fire start-up delay		Stove aesthetics	
		User error		Perceived safety	
		User instruction		Perceived value	
				Taste	

criteria. Field notes and optional photo, audio, and/or video records are also encouraged add context and to help clarify uncertain results.

This structured application of ethnographic methods, as well as the decision to use a quantified Likert-scale survey, may be seen as a compromise between the resources available to most practitioners and the necessity of understanding cross-cultural perspectives and priorities in more detail than is possible with typical engineering research methods. This is in line with Isaac's description of rapid ethnography for industry applications [29], as well as conventional usability tests. The protocol also includes quantitative and objective, as well as qualitative and subjective, measurements in order to increase the amount of relevant information that may be collected from the time spent testing. Both qualitative and quantitative methods are used to evaluate the same criteria, when possible, to add context and allow for comparison and the identification of likely error or bias in the results. The tests are organized into four sections based on the type of methods used. An overview of test sections and methods is shown in Table 2. These tests are meant to be conducted in a kitchen while a cook makes a normal meal to make the results as representative as possible. Abbreviated field and laboratory testing procedures are also outlined in the protocol. These involve applying a subset of the four test sections described in Table 2, and may be of use for preliminary or other work when less detail and context is required.

The test methods and criteria selected for the protocol are meant to be applicable to as many cooking scenarios as possible, with the understanding that each cooking culture and technology will have nuances that are not accounted for. To compensate, methodological shortcomings and missing or irrelevant usability criteria may be identified naturally during by the first few usability tests in a study. Triangulation of the multiple, mixed methods used by the protocol, and especially data collected by interviews and reflective field notes, can offer insights into appropriate changes to improve the effectiveness of the protocol for subsequent testing in a study. In summary, the protocol encourages flexibility and reflexivity in order to increase effectiveness in a given context, but also includes the prescribed set of testing methods and criteria described above as a relevant starting point so that users of the protocol do not need to reinvent the wheel.

Quantitative test results are generally scored from 0 to 4 to align with the International Standards Organization tiers of performance for improved cookstoves (ISO-IWA 11:2012), which uses the same scale for common cookstove testing protocols that cover areas such as emissions, efficiency, and safety, and is familiar to many practitioners [56]. Scores are based primarily on cooks' responses to the survey questions, and are reported alongside objective and quantitative measurements where possible to provide a more complete assessment of the criteria. The protocol spreadsheet calculates averages and a margin of error for quantitative measurement and Liker-scale survey results, and offers a qualitative coding tool for interview results. The margin of error is calculated based on a Student's t distribution given the sample size, standard deviation of the sample, and a confidence level entered by the user.

4. Lira, Uganda field trial

A preliminary evaluation of the proposed cookstove usability protocol was conducted in household kitchens near Lira, Uganda over about three weeks, which is in line with Hartson and Castillo's recommendation of assessing usability protocols in the field [19]. The field study was designed to mimic the length and methods of a realistic usability study as prescribed by the protocol to provide a representative assessment of its effectiveness, with the addition of interviews of the local test administrators and community meetings to allow for additional self-reflection and triangulation of the results. Statistically significant quantitative results were not necessarily expected from the sample sizes allowed by the limited duration of this study; the value in this work was primarily expected to come from the author's experience

Table 2
Protocol test sections and methods.
Source: Moses and MacCarty [14]

	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.

of administering the protocol and expert feedback from the Ugandan test administrators, as per Hartson's recommendations of employing expert, qualitative assessment of a new usability protocol [15].

4.1. Participant selection

Participants with a variety of traditional, improved, and modern stove designs in urban, rural, and institutional kitchens were included in this study in order to use the protocol with as wide a range of cooking technologies and contexts as possible. Rural study participants cooked primarily with wood, which was both collected and purchased. Most rural cooks had used either traditional, built-in-place mud stoves, or three stone fires until ILF introduced an improved clay rocket stove approximately three years prior to this research, which displaced many of the traditional stoves. This presented an opportunity to assess any differences in usability between an improved stove design, which had already been adopted by some cooks, and its traditional counterpart. Cooking was typically done in dedicated thatch-roofed mud brick kitchens. Participants in the town of Lira cooked exclusively with charcoal, which was purchased locally. All participants spoke the Lango language; few spoke English proficiently.

International Lifeline Fund (ILF) assisted with the planning and selection of participants for this field study. ILF has operated a stove manufacturing and distribution program from its office in Lira for about a decade, and employs mostly local staff. They were approached to collaborate in this research due to their relevant experience with cookstove design, ability to provide logistical support, and local expert staff, who were asked to administer the usability protocol. Study participants were selected from communities that ILF had worked with in the past. Most rural participants had been sold subsidized ILF stoves in 2014, and each participant family was given another ILF cookstove and cooking pot as payment for their participation in this study.

In total, 10 households participated in in-home trials of the usability protocol – 8 rural and 2 urban – in addition to 2 institutional kitchens. Testing was done in urban neighborhoods in Lira and the nearby agricultural communities of Ajoodur and Anenober. A description of the kitchens and cooking technologies evaluated is shown in Table 3, and sample photos of the wood and charcoal stoves studied are shown in Figs. 1 and 2, respectively. 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.

4.2. Methods

4.2.1. Protocol administration, administrators, and administrator feedback

The cookstove usability tests were administered according to the protocol as a preliminary field trial. When more than one type of stove was used to prepare one meal, cooks were asked to respond to each survey and interview question for each stove separately, and objective measurements were collected for both stoves simultaneously. Two ILF staff members alternated administering the subjective portion of the tests in each household. As Ugandan women from the Lira region with university educations and experience facilitating Western cookstove programs, they were local experts in terms of culture, cooking, and traditional cookstove design, as well as experts in the improved cookstove implementation programs of ILF. The quality and experience of usability test administrators has been identified as critical to the quality of the results, and their backgrounds and skillsets were key to this study [16]. It should also be noted that while ILF staff to administering the protocol to their own constituents introduces a clear potential for bias, this is likely to be the case for many users of the protocol, who wish to improve their own stoves or programs and will test in communities they have worked with previously. While a study could be designed to minimize these biases with neutral test administrators, for example, this trial was meant to challenge the protocol by not controlling for these biases, as well as to evaluate it in the way it is likely to be applied in practice.

Table 3
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

The lead author was present to assist with the administration of the tests, both by answering questions that arose from ILF staff during testing and to conduct the portions of the tests involving physical measurements to allow ILF staff to focus on the interview and survey portions of the protocol without interruption. He also took field notes, and photographs to more thoroughly document the testing process, in addition to asking clarifying questions of the participant cook (translated through the ILF staff member) when necessary. The lead author has five years of experience in cookstove design, as well, and was therefore able to view the testing process and results through that lens to better judge the likely value to designers, as was suggested by Hornbæk [16]. Two other OSU researchers working nearby in Uganda were also present for one test, each, in a rural household and provided separate field notes and observations of the testing process.

The use of multiple evaluators, as is suggested by Millen [23], also allowed for more than one perspective on usability, which likely increased the variety of new usability problems identified during testing [57] to help validate and improve the protocol more effectively, especially given the limited sample size and time available for the study. The presence of both local and foreign test administrators (ILF staff and the author, respectively) was meant to provide the added benefit of a broader set of expert perspectives on cooking behaviors and usability, as well, in-line with Hartson's recommendation to employ subject-matter experts to evaluate a usability protocol [15].

The data collection form for the tests contains spaces for supplemental notes for most questions and test sections, which the administrators were encouraged to use. These were included in the protocol as a way for test administrators to document context- or household-specific nuances not captured effectively by the tests, and to therefore make it more robust in a wider variety of testing situations in the future. In this field trial, these notes were meant to provide ILF staff with an opportunity to comment on relevance of the test questions and validity of the test methods used, in addition to their perception of the accuracy of participant responses. At the end of the testing period, both ILF administrators were interviewed individually about their impressions of the protocol from their perspectives as local cookstove experts.

4.2.2. Community meetings

A series of community meetings were arranged according to the OSU IRB human subjects protection protocol to inform all study participants of the purpose of the research, gain their consent to participate, and to ask cooks about their impressions of key aspects of the usability of their stoves in a public setting to provide supplemental usability data. These data were then compared with the results of the in-home trials of the protocol as means of triangulating the validity of participants' responses.

4.3. Results

4.3.1. Qualitative usability protocol results

Meaningful qualitative data were produced in the form of field and interview results generated by the ILF test administrators. These were

Fig. 1. WOOD-BURNING HOUSEHOLD COOKSTOVES EVALUATED NEAR LIRA, UGANDA. From left to right: Improved ILF Rural Woodstove, Three Stone Fire, Traditional Mud Stove.
(Photo credit: Nick Moses).

Fig. 2. CHARCOAL BURNING HOUSEHOLD COOKSTOVES EVALUATED IN LIRA, UGANDA. From left to right: ILF Improved Charcoal Stove, Traditional Mud Charcoal Stove.
(Photo credit: Nick Moses).

primarily related to the perceived inconveniences of each stove, such as the limited ash capacity of the improved clay rocket stove or the instability of pots on a three stone fire. Many of these concepts were given as unsolicited additions to survey question responses, which were captured by the ILF test administrators in the included notes sections on the protocol data collection form. In addition, simply being present during cooking and having time to talk with cooks allowed the test administrators and the author to ask clarifying questions whenever an unexpected behavior was observed or response was given. The author's observation during the testing process identified similar areas for improvement in question wording and testing process as were identified by the ILF test administrators, as well, both in the notes of the usability testing form and in the feedback interviews at the end of the study. The amount of time spent with the cook and observation of cooking also allowed for the identification of potentially valuable details beyond usability. These included, for example, direct observation of the relative amount of indoor air pollution present in different kitchens and with different types of stoves and fuels, and insight into the perceived value of a cookstove and motivations for their purchase and use revealed by the detailed conversations with cooks enabled by several hours of in-home testing.

4.3.2. Quantitative usability protocol results

As was expected, insufficient numerical data were collected to provide a statistically significant quantitative assessment of the usability of the stove models tested. The partial exception was the improved clay rocket stove, which was evaluated in five households; the most of any stove tested. A summary of the overall results for the three wood burning stove designs used in rural households, output by the protocol's data processing spreadsheet, is shown in Table 4. These data include the weighted average overall scores for each criterion, which are calculated primarily from the scores and relative importance of each sub-criteria assigned by the cook in the survey portion of the protocol, as well as sample size (n), and highest margin of error (ME) from the contributing sub-criteria for a 95% confidence level (CL). Data are given not to provide a conclusive assessment of usability, but to allow for the discussion of the meaning and validity of the results output by the protocol. Also note that margins of error were not reported for stoves with a sample size of two, as these values would not necessarily be meaningful.

While there was at least one sub-criterion with a relatively high margin of error in each case, the overall category scores were in alignment with survey question results and ILF staff's general understanding of the stove's strengths and weaknesses. This high uncertainty may not necessarily indicate that the score is invalid, but that the results for each sub-criterion should be examined in more detail to determine an effective interpretation of the overall score.

It should be noted that the location-specific needs score was assigned an N/A because the evaluation and scoring process was changed significantly as a result of this field trial. While cooks were asked whether their stoves met the location-specific needs criteria listed in Table 1, the early version of the protocol used only assessed needs that were important to a cook, but did not effectively distinguish between needs that were not important to a cook, and needs that were important but not met by a stove. The data collected were insufficient to rate the stove according to the revised protocol, which accounts for both the importance of a criterion and whether or not a stove meets it, and were therefore discarded. The observation and interview portions of the tests succeeded in eliciting relevant location-specific needs information even in the short timeframe of this study, however, such as the previously

Table 4
Overall usability results: wood burning household stoves.

		Improved Clay Rocket Stove		Three Stone Fire		Mud Stove	
		n = 5		n = 2		n = 2	
		Score	Highest sub-cat. ME (95% CL)	Score	Highest sub-cat. ME (95% CL)	Score	Highest sub-cat. ME (95% CL)
I.	Fuel cost and convenience	2.2	1.89	1.5	–	1.0	–
II.	Cooking performance	3.0	2.72	1.3	–	2.2	–
III.	Operability	2.2	0.68	2.9	–	1.8	–
IV.	Maintenance	2.9	2.07	4.0	–	2.6	–
V.	Comfort	3.2	0.68	2.3	–	2.6	–
VI.	Location-specific needs	N/A		N/A		N/A	

Table 5
Usability sub-criteria results: wood burning household stoves.

I.	Fuel cost and convenience	Improved Clay Rocket Stove			Three Stone Fire			Mud Stove		
		n = 5			n = 2			n = 2		
		Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)
A.	Fuel availability	3.0	2.8	*1.26*	1.0	2.3	–	1.0	2.5	–
B.	Fuel preparation time	1.0	1.7	*1.89*	2.0	2.0	–	1.0	3.0	–
Overall performance tier:				2.2			1.5			1.0
Highest subcategory ME:				1.89			–			–

II.	Cooking performance	Improved Clay Rocket Stove			Three Stone Fire			Mud Stove		
		n = 5			n = 2			n = 2		
		Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)
A.	Cooking speed	3.8	4.0	*0.56*	1.0	2.5	–	3.0	3.5	–
	Cooking duration (measured - hh:mm)	2:59		*1:05*	2:23		–	2:27		–
B.	Firepower control	2.2	3.8	*2.04*	0.5	2.5	–	2.0	4.0	–
C.	Firepower range	1.6	2.4	*2.72*	0.5	3.5	–	0.5	2.5	–
D.	Use of all pots and pans	3.8	4.0	*0.56*	4.0	2.0	–	3.0	2.5	–
Overall performance tier:				3.0			1.3			2.2
Highest subcategory ME:				2.72			–			–

III.	Operability	Improved Clay Rocket Stove			Three Stone Fire			Mud Stove		
		n = 5			n = 2			n = 2		
		Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)
A.	Fuel feed entry size	2.0	4.0	*0.00*	4.0	4.0	–	2.0	4.0	–
B.	Tending frequency (measured)	4 min.		*1.8 min.*	6 min.		–	5 min.		–
	Reloading frequency (measured)	43 min.		*15.7 min.*	57 min.		–	50 min.		–
C.	Tending/refueling effort	3.5	3.8	*0.68*	1.3	2.0	–	2.7	2.0	–
D.	Visibility of fire	2.5	4.0	*0.00*	3.8	4.0	–	2.5	4.0	–
E.	Ease of lighting	2.0	4.0	*0.00*	3.0	4.0	–	1.0	4.0	–
F.	Fire start-up delay	0.0	4.0	*0.01*	0.0	2.0	–	0.0	4.0	–
G.	User error	3.0	4.0	*0.40*	3.0	4.0	–	3.0	4.0	–
Overall performance tier:				2.2			2.9			1.8
Highest subcategory ME:				0.68			–			–

IV.	Maintenance	Improved Clay Rocket Stove			Three Stone Fire			Mud Stove		
		n = 5			n = 2			n = 2		
		Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)
A.	Routine maintenance	**2.0**	4.0	*2.07*	**4.0**	3.0	–	**1.0**	3.5	–
B.	Long-term maintenance	**3.7**	4.0	*0.56*	**4.0**	3.5	–	**4.0**	4.0	–
Overall performance tier:				**2.9**			**4.0**			**2.6**
Highest subcategory ME:				2.07			–			–

V.	Comfort	Improved Clay Rocket Stove			Three Stone Fire			Mud Stove		
		n = 5			n = 2			n = 2		
		Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)	Score	Relative Weight	ME (95% CL)
A.	Perceived safety	**4.0**	4.0	*0.00*	**0.0**	2.0	–	**2.0**	3.0	–
B.	Perceived smoke exposure	**2.6**	3.0	*0.68*	**0.0**	4.0	–	**2.5**	2.5	–
C.	Cooking area soot deposits	**0.0**	2.0	*0.00*	**0.0**	2.0	–	**0.0**	2.0	–
D.	Pot soot deposits	**0.0**	2.0	*0.00*	**0.0**	2.0	–	**1.3**	2.0	–
E.	Cooking height	**4.0**	4.0	*0.00*	**4.0**	4.0	–	**3.5**	3.5	–
F.	Stove aesthetics	**3.6**	3.6	*0.68*	**3.0**	3.0	–	**3.5**	4.0	–
G.	Perceived durability	**4.0**	4.0	*0.00*	**4.0**	4.0	–	**2.0**	4.0	–
H.	Perceived value	**3.8**	4.0	*0.56*	**4.0**	4.0	–	**4.0**	4.0	–
I.	Taste	**3.6**	N/A	*0.68*	**1.5**	N/A	–	**3.5**	N/A	–
Overall performance tier:				**3.2**			**2.3**			**2.6**
Highest subcategory ME:				0.68			–			–

unknown function of drying firewood above a cooking fire valued by some cooks, as well as the high variance in the importance of needs such as portability and drying food from household to household.

The more detailed sub-criteria results output by the data processing spreadsheet are shown in shown in Table 5 with the exception of the location-specific needs scores, as explained above. The margins of error were relatively small for some sub-criteria, such as “cooking speed” and “use of all pots and pans” categories under the “Cooking Performance” criterion, indicating a strong agreement between all 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 other community members not included in the study. The wide margins of error for other criteria could indicate that there is either high variation in opinion between cooks, that the test results for these criteria were not valid, or that the sample size was insufficient to elicit a reasonable average value.

For the sake of this analysis, a margin of error of 0.5 points was used as the upper limit to indicate the quality of the subjective criteria scores, which are rated from 0 to 4. Higher or lower margins of error may be acceptable for different testing purposes, however, and may be correlated to adoption, usage, and other metrics in future research. With a 95% confidence level, the results for the improved clay rocket stove ($n = 5$) yielded a margin of error of less than 0.5 points for 10 out of 23 of the sub-criteria evaluated (43% of sub-criteria). Hypothetically, if the sample in this study was generally representative of the larger community, 10 tests would have increased the percentage of sub-criteria with margins of error below 0.5 points to approximately 75%, and 15 tests would have increased this figure to about 83% of criteria (and if this sample was not representative of the larger population, these percentages would be even smaller). This indicates that while a larger sample size may have increased the accuracy of some criteria, as well as the validity in extrapolating usability results to the larger study population, other criteria would likely still be uncertain.

4.3.3. Community meeting results

Most responses in a community setting were positive towards improved ILF cookstoves compared to traditional stoves, though also included discussions of design shortcomings. Participants identified the small fuel feed of the ILF wood rocket stove as a nuisance compared to local stove options, and claimed that the newest ILF charcoal stove was not durable enough. All responses received were in line with usability criteria already incorporated into the protocol, or issues related to cost and value that were determined to be outside the scope of the usability protocol.

4.3.4. Test administrator feedback

ILF test administrators suggested substantial improvements to protocol during the first few in-home trials, which were incorporated into subsequent tests. These changes included rewording several questions to improve clarity, adding spaces for field notes after each question or group of questions, and administering the survey and interview portions of the test one after another in a seamless, conversational format. As the testing process became more streamlined and conversational during the course of the study, the participants spoke more naturally and more often, potentially producing higher quality responses.

Additional informal interviews with both test administrators at the end of the study confirmed that, based on their past experience working with traditional and improved cookstoves in these communities, despite the need for improvements the content of the protocol was generally appropriate and the results obtained were in-line with their expectations and understanding of local cooking. They also commented that while they had surveyed these communities before, they had not done in-depth testing in individual homes. They said that this approach added value, but also took significantly more time per household. This feedback led to the development an alternate field testing procedure to give test administrators the option to collect basic usability information from cooks in the span of a few minutes without observing a cooking event, as opposed to the several hours required by the full protocol.

4.4. Discussion

4.4.1. Interpretation of quantitative and qualitative results

Portions of the quantitative results from the trial of the protocol in Uganda were likely relevant and meaningful. For example, the three stone fire scored a perfect 4 for the overall maintenance criteria, which is in line with expectations for a free, virtually indestructible cooking technology. In at least one case, the validity of a numerical result was not just uncertain but doubtful. Cooking speed for the ILF improved clay rocket stove was assigned a higher score than were the three stone fire and mud stove. Measurements indicated that while it may save fuel, it was not, in fact, noticeably faster. This contradiction could be due to an erroneous belief amongst cooks due to past ILF training and education, or could be a case of cooks telling ILF staff what they thought they wanted to hear in order to remain in their favor for future projects, etc. In either case, the use of multiple methods allowed for this uncertain result to be identified.

Larger sample sizes would have produced numerical results with smaller margins of error. The Cookstove Field Study Resources guide [58], published as a part of the upcoming International Standards Organization standard for cookstove testing (ISO/TC 285) suggests a minimum sample size of 15 for uncontrolled field research, which may be a more reasonable starting point for future usability research [59]. High variation for some criteria, however, such as routine maintenance requirements or perceived fuel availability, appear to be a function of real differences between individuals. These may be explained by differences in household income levels, individual standards for stove aesthetics and upkeep, or perhaps variation in comfort with negative responses, where some cooks responded in a way that would please ILF staff and others responded with more direct opinions. In addition, the results for these criteria may not have truly normal distributions, but follow bimodal or other patterns, and are likely poor candidates for purely quantitative assessment and basic statistical tests.

These uncertain results, indicated either by low statistical significance or contradictory results from subjective and objective test methods, may be seen as indicators that criteria must be assessed in more detail to be adequately understood. This is in-line with the work of Malakar and Jurisoo, Lambe, and Osborne, who suggest that a high level of detail is needed to understand users' decision-making [39,43]. Even with large sample sizes of 30 or more, as is recommended by common technical cookstove tests like the Kitchen Performance Test [60], it is possible that though apparently statistically significant results could be attained, they would in effect be hiding more complex realities. For example, if some cooks in a community collect wood while others purchase it, there are likely to be two distinct (bimodal) trends in results related to fuel availability. Both sets of users have different needs and perspectives, and neither would be served by an average value of the entire population's perceptions, regardless of statistical significance.

In cases like these, qualitative data, either taken during the tests or during follow-up research, would be necessary to understand cooks' differing perspectives, any implications for stove design or selection, and to design possible improvements to quantitative testing methods to capture these phenomena more effectively. Frequent data analysis in between tests during a study may allow for the identification of such criteria, which can then be targeted with qualitative interview questions during subsequent tests to achieve a more complete understanding without necessitating follow-up work. It should also be noted that this type of qualitative work and the resulting understanding of local perspectives and needs could be used to increase the construct validity of the protocol; more specific quantitative survey questions or measurements could be created to evaluate these findings in future testing in the same context. A separate, qualitative preliminary study could also be used to adapt the protocol to a specific context and maximize the quality of the information gained from the resources invested in testing. In addition to increasing the validity of quantitative results, this could reduce the need for qualitative data and interpretation to clarify uncertainty.

While there was significant variation in the numerical results, the qualitative data collected were sufficient to be valuable to a designer in the author's experience as an engineer and cookstove designer. Even for the mud stove and open fire that were included in just two tests, each, this work amounted to a significant amount of time with participants; several hours' worth of observation, note-taking, and conversation. Some of the data collected, such as the instability of pots on a three stone fire, overlapped with quantitative survey results. Others, such as the limited ash capacity of the improved clay rocket stove, were new to the author and not already included in the survey or other portions of the protocol.

This information provided the most direct appraisal of user's wishes for potential design changes, and also appeared to be a more efficient way to gauge the causes of variation in responses from cook to cook than the collection of more quantitative data from additional repetitions of the test. Especially in the preliminary stages of a cookstove design, dissemination, or research program, the collection of qualitative and anecdotal quantitative data from a small number of tests may be a more efficient use of resources than a larger study designed to achieve high statistical significance. Approximately three tests on the improved clay rocket stove were needed before the majority of qualitative usability insights were collected, though diminishing amounts of new data were elicited on each subsequent test. It should be noted that qualitative data and observations pertaining to user perspectives and priorities, purchasing or adoption decisions, and other factors relevant to cookstove programs and development work beyond usability did not approach saturation during this study. Engagement with cooks using applicable methods, including ethnographic methods, for a significantly longer study is likely to continue to elicit valuable information that may inform product or program design.

It should also be noted that spending several hours at individual homes, and several days or longer in the same village or region, elicited information that would not have been available during shorter surveys more commonly used in cookstove work. In addition to giving community members time to adjust to the presence of researchers, the approach used in this study gives stove designers the opportunity to carefully observe the cooking practices of multiple cooks, which can be an invaluable way to create first-hand understanding of user needs and therefore support effective engineering design work, based on the author's experience. The amount of time spent in-person, in kitchens, when tests are repeated for several days or weeks, gives this protocol an ethnographic element that is absent from shorter survey- or interview-only user research, for example. It should be noted that this approach does not produce the depth of understanding that could be gained by a traditional ethnographic enquiry spanning months or years, however. The limitations discovered in this research in fact highlight the additional value that could be provided by more thorough research methods, such as ethnography.

The amount of time spent in the community also allowed the researchers to collect relevant information outside the bounds of usability and user needs that could be critical to the overall success of a cookstove program, such as the structural and cultural factors suggested by, Davies, Otte, Wang, and Malakar, Grieg, and van de Flirt, and others [37,40,41,44]. For example, the two participant households with the most apparent wealth (in form of brick houses with steel roofs, as opposed to thatched-roof mud huts) were the only two visited that used three-stone fires. During the interviews, it was revealed that neither cook was as concerned about the cost of firewood as their neighbors, so was less incentivized to purchase a fuel-efficient stove though they clearly could have if they placed a high enough value on other potential benefits, such as improved air quality. While not directly related to usability, these kinds of insights are not always collected through other testing protocols or in the course of a cookstove implementation project, yet could prove important to the success and impact of a stove design or program. As this is the same ultimate goal of the usability protocol, the ability to elicit additional relevant information should be noted as a valuable benefit of the in-home usability testing and the qualitative and ethnographic methods used by the protocol.

4.4.2. General discussion regarding validity

The ILF test administrators and the lead author perceived that the protocol was effective in eliciting generally relevant, if mostly qualitative, usability data from the participants and stove designs included in this study, despite the shortcomings described above. The protocol was also applicable to disparate technologies (i.e. wood-burning and gas stoves), and elicited usability insights that were not present in the literature and were new to the author, such as the hesitance to use otherwise highly usable LPG gas stoves in Uganda due to the risk of explosion. In the case of commercial stoves, many test questions regarding personal preference and cultural norms appeared to be less relevant, which might be expected from an employee who makes fewer decisions about their cooking technology and practices, and may prioritize speed and convenience more highly than a cook feeding their own family.

Results from community meetings generally mirrored in-home usability testing results, indicating that participants were likely responding relatively honestly and directly in both cases. Attitudes towards ILF stoves tended to be more positive in the community setting, however, and more ambivalent in individual tests, suggesting that while cooks were considering the strengths and weaknesses of each stove model to an extent, they were less likely to provide negative opinions in a public setting. This may have been out of politeness, embarrassment, in order to receive future goods and services from ILF, or for other reasons, though this indicates that in-home testing likely provided a more relevant and in-depth assessment of cooks' perceptions. The concerns raised at community meetings related to value proposition raised are also valid factors in purchasing and adoption decisions, and deserve consideration in future research.

4.4.3. ILF's relationship with the participants

The pre-existing relationship between ILF and the participants – specifically as giver and receiver of subsidized goods – created a clear potential for bias in the usability testing results. The authors recognize that participants may have altered their responses, been more engaged or willing to participate, changed the type and quality of the meals they cooked, etc. However, personal relationships between ILF staff and participants were also leveraged to simplify study planning and, more importantly, increase the participants' comfort with the researchers' presence, which potentially increased the odds of open and honest responses. More testing with participants and test administrators without this prior connection would be necessary to explore the effects of this issue on usability testing results, which likely varies in different cultures and depending on the nature of the prior relationship.

The goal of this study was to evaluate the protocol, not the cookstoves tested, however, and the experience of the ILF administrators with the individuals, communities, and cookstoves included in the study made them well-qualified to judge the effectiveness of the protocol in this context. Their identities as women and as local experts, intimately familiar with the culture, context, and cooking practices of Northern Uganda, likely allowed them to relate and communicate with participants exceptionally well, while their formal education and experience as employees of a Western NGO allowed them to support the research goals of the authors, as well as to communicate their impressions of the protocol effectively. Due to these benefits, local experts with similar backgrounds may make for high quality test administrators in future usability studies even if they have a prior connection to study participants, though the comparative effectiveness of test administrators with a third-party positionality should be evaluated in future research.

4.4.4. Other sources of error

Several complicating factors and potential sources of error within the testing process were also identified during test administration and observation. At times children assisted mothers with cooking tasks outside the kitchen, making the complete observation of the entire cooking process impossible. Cooks also frequently waited for the research team to arrive to begin cooking, which may have impacted their cooking habits and especially stove lighting practices; if a stove was lit in the morning to make breakfast or tea, it may not have normally been re-lit to cook the mid-day meal, as was often observed. Similarly, when multiple stoves were used at the same time, they may have each been used to their strengths, just as various kitchen appliances in high-income countries are used for specific cooking tasks. In this case, the usability results for each stove may have been different than if they were used in isolation, though perhaps still valid in the actual "stove stacking" arrangement used by the cook. The author also suspects that more elaborate meals were sometimes cooked for the research team. While not as representative of a typical daily meal, these additions did give the research team opportunities to observe cooking practices for a greater range of dishes, along with potential usability implications.

It also should be noted that throughout the testing process and in discussions with ILF staff, additional variables were identified that could impact testing in other scenarios in Uganda or elsewhere. These include varying levels of cooking experience with a stove, seasonal changes in cooking habits due to agricultural or weather patterns, and varying cultural norms that may impact a participant's responses. For example, ILF staff suggested that even in other regions of Uganda people are less inclined to give direct responses, but instead are more likely to try to please a test administrator with their answers. The authors expect that there are a wide range of potential complications that will affect testing in different countries and regions which will need to be identified and addressed by study designers and test administrators to maximize the quality of usability testing results.

The identities of the researchers present during testing also played a clear role in the cooks' behavior, and potentially in their responses. With few exceptions, men besides the author were not present in the kitchen during testing. The impact of gender became apparent during one test where two female Oregon State University researchers were present in addition to the male author to observe. The atmosphere in the kitchen was less formal than in other tests, and this was the only case in which the cook made significant attempts to engage with the researchers. It is not clear from this work how gender may have impacted testing results, though in this context (and in many others with similar gender roles) it is likely that the presence of men in the kitchen is uncomfortable for the cook and may reduce engagement, as well as possibly alter responses.

These potential sources of error are to an extent inevitable in any testing involving human subject research; especially in a field environment. Though larger, more carefully controlled, and more expensive studies could reduce their impact, many practitioners who wish to conduct usability testing may not have the resources for studies significantly larger than the one described here and are likely experience comparable levels of bias in their work. With this in mind, discussion of these potential issues has been included in protocol as a result of the study in Lira so that future test administrators may at least be aware of and account for them as much as is possible, and therefore improve their ability to interpret test results.

Similarly, the presence of biases does not mean that usability testing cannot be worthwhile, even with small sample sizes. Valuable, if primarily qualitative and anecdotal, usability data was still obtained in this study despite the small sample sizes, and the quantitative data provided a means of triangulating and understanding results and likely errors, even though much of it was not able to stand on its own as valid, statistically significant results. The authors also hope that a positive, value-added experience with user research and qualitative and ethnographic methods may also encourage practitioners to invest more effort in more thorough follow-up usability studies or on similar approaches in other areas of their work, as have been identified by several researchers as necessary to achieve a thorough understanding of user preferences and decision-making [40,41,44].

5. Guatemala field trial

A second trial of the protocol was conducted in four peri-urban households near Antigua, Guatemala to provide an assessment in a different region with a different cooking culture and technologies. The protocol, with the revisions described above from the study in Lira related to question wording, survey procedure, and location-specific needs questions, was administered by student researchers without the presence of expert local interpreters or the author. This provided an opportunity for a trial in a more challenging and realistic use case without outside assistance.

5.1. Participant selection

Participants used both improved stoves and traditional open fires during testing, either with a plancha (griddle) cooking surface or a stock pot. StoveTeam International and the local EcoComal cookstove factory staff assisted with the selection of participants from communities they had worked with in the past. Study participants were aware that the researchers were affiliated with StoveTeam International, the US-based provider of the improved, wood burning plancha stoves sold at a subsidized price to the people in the area. The interviewed cooks were providing feedback on the stove designed and manufactured by the local EcoComal factory in collaboration with StoveTeam. They may or may not have used that or a similar stove at home in the past. Examples of the traditional and EcoComal cookstoves are shown in Fig. 3.

5.2. Methods

The usability protocol was administered by OSU student researchers fluent in Spanish without the assistance of interpreters or prior instruction from the author. In addition to applying the protocol, the researchers summarized the usability findings, as well as their impressions of and suggestions for improvements to the protocol in a separate report.

5.3. Results

Key outcome of this work included the identification of shortcomings in the test procedures and test administration instructions, as well as opportunities for improvements to testing and data collection procedures. Several objective tests and survey questions which did not pose problems in Uganda required clarification in Guatemala. Areas of the instructions were also identified as unclear or inadequate for the protocol to function as a self-contained document in the short timeframe of this study without the benefit of the guidance from the author that was available in Uganda.

In addition, the value of using two test administrators in order to allow one to focus on conversation with participants without interruption, as was done in Lira, was by confirmed the test administrators in Guatemala. New ways to apply the survey questions were also suggested. These include asking cooks separately about their current stove and an ideal stove to prevent confusion when weighting the relative importance of criteria, as well tablet-based electronic data collection tools. As with the trial of the protocol in Lira, some usability criteria returned consistent and plausible results, while others showed high variation between participants or evidence of contradictory or biased responses. A detailed analysis of the usability data is not presented here, as it would not add to the discussion already included on the assessment of the larger dataset from Lira.

Fig. 3. COOKSTOVES EVALUATED IN GUTEMALA. Traditional three-stone fire with plancha (left) and Stove Teams International plancha stove (right). (Photo credit: Josephine R. Crofoot).

5.4. Discussion

This work identified opportunities for improvement to the usability of the protocol itself as a self-contained resource for test administrators, and provided another set of perspectives on its use and ways to improve the test methods and questions. The trial of the protocol in Central America also demonstrated that while the methods and criteria were generally applicable to a different cooking culture and set of technologies than were evaluated in Uganda, new shortcomings became apparent in this new context. This is likely to be the case in any location, and highlights the need for reflexive testing practices that identify and account for these issues, or qualitative preliminary research to adapt the protocol to a new context in advance. These insights will be incorporated into a future version of the protocol.

6. Preliminary assessment of a third-party trial near Gulu, Uganda

Usability tests, again with the updated protocol resulting from the Lira study, were conducted by Aid Africa with 20 users of their wood rocket stove (shown in Fig. 4) near Gulu, Uganda. This work was done to provide an assessment Aid Africa's cookstove which has been distributed free of charge in these communities. Their results and feedback were shared with the authors for assistance with the interpretation of the results, as well as to offer another opportunity to validate the protocol and to understand its limitations. The results and discussion presented here are limited to feedback from the test administrators and Aid Africa staff, as well as the authors' assessment of the application of the protocol and a brief overview of the results. Usability testing data are not reported in detail to avoid duplication of the discussion of the Lira study, as well as to avoid any premature conclusions before the results can be fully analyzed and discussed with Aid Africa.

Fig. 4. AID AFRICA ROCKET STOVE EVALUATED NEAR GULU, UGANDA. (Photo credit: Aid Africa).

6.1. Methods and participant selection

Tests were administered by four Ugandan women – two university student volunteers and two Aid Africa staff – with no foreigners present. Cooks were approached to participate on a daily basis without advance notice so that the they would be less likely to alter the stove they used or cooking practices on account of the testing. The administrators reviewed their results with the Aid Africa director each evening after testing to work through any uncertainty in the testing process. Neither the director nor the administrators received instruction from the authors beyond what is included in the text of the protocol.

6.2. Results

The test administrators reported initial difficulty in understanding some aspects of the protocol, but that the testing process became clearer after a few days of testing. This is corroborated by the test results, which show several consistent administrator errors from the first two days of testing, then few or no errors on subsequent tests. Objective tests, such as the visibility of the fire, were also scored consistently across all 20 tests by each of the 4 administrators.

Regarding the effects of sample size on statistical significance of the quantitative usability scores, the 20 tests conducted produced results with a margin of error of less than 0.5 points for 18 of 24 usability sub-criteria evaluated (on the 0–4 scoring scale, and with a confidence level of 95%). This amounts to 75% of sub-criteria evaluated, as opposed to the 43% of sub-criteria with the same maximum margin of error from the sample of 5 ILF rocket stoves tested in Lira. Similarly, changes made to the protocol after the Lira study, such as improvements to the location-specific needs survey questions, appeared to function as intended.

6.3. Discussion

The feedback from Aid Africa and preliminary assessment of the test results demonstrated that while the protocol can be used and applied effectively from just the included instructions, there is room for improvement to reduce the learning curve for test administrators. This mirrors the shortcomings found during the administration of the four tests in Guatemala. Specific opportunities for improvements to the instructions identified through this study will also be incorporated into a future version of the usability protocol.

In addition, this work reinforces the assessment of the effect of sample size on statistical significance found in the results from Lira; a larger sample size increases the number of statistically significant criteria scores, though some of them are still not likely to produce statistically significant results and would require qualitative follow-up for clarification. The final assessment of this data will include an investigation of the effects of local versus foreign test administrators, Aid Africa staff test administrators versus volunteer administrators, the sampling method used, and a comparison to any additional testing done with cooks using baseline, traditional stoves.

7. Summary of field trials in Uganda and Guatemala

An overview and comparison of the three field trials described in this article is presented below. Results are described in terms of impact on the refinement and validation of the protocol. Usability measurements from each study are not included, as they would not necessarily provide a fair comparison of usability across different technologies and in different contexts (Table 6).

Table 6
Field trial summary.

	Lira, Uganda	Antigua, Guatemala	Gulu, Uganda
Sample size	10	4	20
Description of stoves tested	Traditional and improved wood and charcoal stoves; two commercial stoves	Traditional and improved wood-burning plancha stoves	Improved wood rocket stoves
Description of test administrators	Staff from local improved cookstove supplier	US university students	Ugandan university students and staff from local improved cookstove supplier
Key outcomes	• Test procedure and content improved • Preliminary, positive feedback on effectiveness of the protocol from local experts • Implications of small sample sizes tested	• Opportunities for improvements to test procedure and administration instructions identified	• Opportunities for improvements to test administration instructions identified • Implications of moderate sample sizes tested • Validation of improvements to procedure and content made after the Lira study

8. Laboratory trial

The *Cookstove Characteristics Evaluation* portion of the protocol, which includes physical measurements and objective observations of stoves (described in Table 2), was conducted on stoves at the Aprovecho Research Center in Cottage Grove, Oregon, USA, a cookstove design and testing organization that has developed stove technologies for several decades. These tests were done to help refine and calibrate this portion of the test on a wide, if not all-inclusive, range of improved cookstove designs, as well as to understand the likely value of laboratory versus field testing.

8.1. Cookstove selection

A subset of 11 stove designs was chosen from the improved and local stove designs from around the world housed at the Aprovecho Research Center. Stoves were selected to represent models that are used widely, as well as to represent a variety of fuels and geographical regions. These included five wood-burning rocket stoves, four top lit updraft pellet stoves (TLUD's) – two with electric fans – and two charcoal stoves, which were taken from India, Southeast Asia, Africa, and Latin America. This sample was deemed sufficient by the author and Aprovecho staff to be representative of most common stove types, though it should be noted that local, artisanal stove designs were likely underrepresented, as more manufactured stoves have been brought to the Aprovecho Research Center over the years by their designers and promoters.

8.2. Methods

The methods prescribed by the *Cookstove Characteristics Evaluation* section of protocol were used to conduct these tests. These include measurements of the fuel feed entrance of the stove, a count of the number of burners, and objective observations of the exhaust venting design, fuel type used, and materials and methods required for lighting. The exact testing protocol followed is available online [14].

8.3. Results

This work resulted in the calibration of the fuel feed entrance size criteria values, expansion and clarification of the language included in this portion of the protocol to accommodate the various features of the stoves tested, as well as a verification that the measurements included were relevant to common biomass stove designs. A list of stove models tested and the results from each is shown in Table 7.

The results of these tests are used in the calculation of several usability sub-criteria and could be input into the Excel spreadsheet to provide a partial set of results. Without the scores produced by the user

Table 7
Laboratory testing results.

Stove Description		Manufacturer	Fuel feed cross-sectional area (cm^2)	Ease of lighting		Indoor soot evaluation		Burner count
				Kindling or special lighting materials are required	Fire must be lit in an enclosed space within the stove	Gas, alcohol, or electric power	Sealed chimney	
1.	Wood rocket	Greenway	161	Yes	Yes	No	No	1
2.	Wood rocket	Stove Tec	84	Yes	Yes	No	No	1
3.	Wood rocket	Potential Energy	90	Yes	Yes	No	No	1
4.	Wood rocket	Envirofit	96	Yes	Yes	No	No	1
5.	Wood rocket	Biolite	114	Yes	No	No	No	2
6.	TLUD	Alpha Renewable Energy	189	Yes	No	No	No	1
7.	TLUD	Mini Moto	71	Yes	No	No	No	1
8.	TLUD with fan	Oorja	79	Yes	No	No	No	1
9.	TLUD with fan	Philips	95	Yes	No	No	No	1
10.	Charcoal	Burn Mfg.	N/A	Yes	No	No	No	1
11.	Charcoal	Gyapa	N/A	Yes	No	No	No	1

survey portion of the protocol, however, there is not enough data to utilize the scoring system of the protocol as intended. These data are reported directly, here, instead, as they may be of more value individually as a way to compare the features and relative usability of multiple stoves, or to provide a preliminary, low-cost way to predict elements of usability related to lighting, tending, soot deposits, and the number of pots or pans that can be used simultaneously. It should be noted that 6 individual metrics could be collected for each stove, compared to about 50 from each household during field testing.

8.4. Discussion

While these results were useful in calibrating and validating the measurements included in the protocol, contextual information about the meaning of a given measurement was lacking, and with no personal experience using some of the stove designs tested it was difficult to correlate the likely positive or negative qualities perceived by a cook to physical measurements and observations. Stove measurements performed by a test administrator with more relevant experience in a given region and context, or the simulation of a cooking event in the laboratory by a representative cook or researcher, may provide additional value, though are also likely to be of limited applicability to real-world usage compared to field testing.

Of note is the comparisons of rocket stoves designed primarily by Western organizations to a stove designed and sold in India (the Greenway rocket stove), which demonstrates that the Indian stove had significantly larger fuel feed entrance. This design decision typically comes at the expense of technical performance metrics, such as thermal efficiency and improved emissions, but may enhance usability of the stoves in terms of the ease of tending, reloading, and visibility of the fire. It may be that the designers of this stove are closer to and more familiar with the contexts of their users due to geography or economic necessity, if they rely on customers for direct sales, and therefore tend to prioritize usability more than Western designers. This insight may support the hypothesis that humanitarian stove designers tend to neglect usability in their products.

9. Conclusion

This research demonstrated that the proposed usability protocol is capable of generating valuable information about cookstove usability, and to an extent situate it in the larger context of cooks' priorities. The authors hope that this protocol and the research described here will be of immediate use to cookstove practitioners as a guide for increasing their understanding of usability and balancing user, technical, and other design goals. More research is needed, however, to demonstrate the protocol's effectiveness in enabling improved stove design and program decisions, as well as the relationship between usability scores, stove usage, and overall impact. The methods used to assess quantitative results in particular, as well as their validity, require further assessment.

These trials of the protocol led to the identification of multiple opportunities for improvements. Survey question language and procedure, which has been made conversational instead of formulaic, as well as the location-specific stove needs section, were improved as a result of the study in Lira, Uganda. Alternative, abbreviated testing procedures were also developed. The work in Guatemala resulted in the identification of opportunities for improvements to the usability of the protocol, itself, for test administrators, as well as validation that the methods and criteria included were largely applicable in a different culture and with different cooking technologies. A preliminary assessment of the testing done near Gulu, Uganda, highlighted additional areas where the instructions for test administrators could be improved, as well. Improvements identified as a result of the work in Lira have already been incorporated into the protocol, and those identified in Guatemala and Gulu will be included in a future revision.

This research also led to insights about strengths and weaknesses of various study design choices and ways to administer the protocol. The study in Lira demonstrated the value of local expert test administrators who can accurately interpret user responses and put results in context, as well as the potential for bias introduced by using non-neutral parties as test administrators. In addition, the value of using two test administrators in order to allow one to focus on conversation with participants without interruption was identified in Lira and confirmed in Guatemala. Finally, the trial of the protocol in the laboratory demonstrated that while the physical measurements included apply to a wide range of designs, limited information can be collected without the context provided by field testing, or at least a test administrator familiar with cooking practices for a stove in a given context.

Based on this work, a preliminary understanding of the value of the qualitative and quantitative test results has also been developed. Qualitative results may provide valuable information with approximately three test repetitions per stove model, though reasonable quantitative results for the majority of usability criteria require larger sample sizes than were included in the study in Lira and Guatemala; likely 15 or more. More research is needed to establish sample size recommendations, though a preliminary assessment of the study in Gulu confirmed the need for larger sample sizes to produce significant quantitative results, as well as the need for additional qualitative methods to evaluate some criteria effectively even with larger sample sizes. The calculation of the margin of error for each criterion allowed for the identification of criteria that could be assessed effectively with comparatively simple quantitative data, as well as criteria that would have required more thorough qualitative assessment to be adequately understood. This feature should allow for a more efficient use of both quantitative and qualitative methods in future testing.

In addition, since personal preferences, practices, and needs related to some usability criteria were found to differ significantly between households in the Lira study, just as they do within communities in high-income countries, cookstove designers and implementers may expect that in most cases it will take more than one stove design or program approach to meet the needs of an entire community. The process of spending several hours with, and systematically observing, a cook's habits may also offer new and valuable insights to cookstove practitioners related to usability and other aspects of their work. While this level of engagement is brief compared to traditional applications of ethnographic methods, it may be seen as a compromise between the time and resources available to most cookstove practitioners and the quality and breadth of the insights they may gain. More thorough and time-intensive research methods, such as traditional ethnography, would likely produce significantly more detailed and complete data than the proposed usability protocol.

Future work to further validate and improve the protocol could include trials with upcoming revisions resulting from the Guatemala and Gulu studies, trials in other locations and contexts and with cooking technologies not included in this study, as well as the specialization of the methods and criteria for a specific context or cooking technology; this protocol may serve as a starting point for those who wish to specialize or adapt it to better suit their needs. Research could also be designed to correlate usability testing results to cookstove purchasing or adoption, sustained usage, and health and environmental impacts. Feedback from usability studies directly included in stove or program design may also be used to inform future improvements to the protocol. Several organizations that design, test, or implement cookstoves are currently evaluating the protocol, and their feedback will be incorporated into a future revision.

Elements of this work are incorporated into the International Standards Organization ISO/TC 285 standard for cookstove testing, currently under development, to improve awareness of the need for usable designs. The usability protocol may be referenced directly by the completed standard, as well [59]. The full protocol will be published on the Global Alliance for Clean Cookstoves website, and is also available online for use and evaluation from the Oregon State University website [14]

This effort to promote and increase the standardization of cookstove usability testing will also hopefully lead to greater accountability for the design and selection of effective, appropriate stoves for a given context, as well as facilitate effective communication about usable design strategies through increased awareness of usability and related terminology. More broadly, the interdisciplinary approach used in this research, as well as in the design of the protocol, may serve as a model for future humanitarian engineering work, where traditional engineering skill-sets alone are not necessarily sufficient for effective cross-cultural design to adequately address complex problems.

Acknowledgements

The authors would like to thank International Lifeline Fund's US and Uganda-based staff for facilitating the field trial of the protocol in Lira – especially Rebecca Apicha and Jennifer Auma for administering the tests – as well as Sam Bentson and the Aprovecho Research Center for making their collection of stove models available to us. We would also like Dr. Shaoezeng Zhang for his guidance in this research, and fellow OSU researchers Grace Burleson and Jen Ventrella for observing and providing feedback on the usability testing process in Uganda. We are also grateful to the individuals and organizations currently evaluating and providing feedback on the protocol to help us make improvements, especially Peter Keller and Josephine Crofoot for their work trialing the protocol and providing feedback for this article. 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] WHO, Household Air Pollution and Health, [Fact sheet] (2016) http://www.who.int/mediacentre/factsheets/fs292/en/.
[2] T.C. Bond, E. Bhardwaj, R. Dong, R. Jogani, S. Jung, C. Roden, D.G. Streets, N.M. Trautmann, Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850–2000, Global Biogeochem. Cycles 21 (June (2)) (2007).
[3] V. Ramanathan, G. Carmichael, Global and regional climate changes due to black carbon, Nat. Geosci. 1 (4) (2008) 221–227.
[4] V.R. Putti, M. Tsan, S. Mehta, K. Srilata, The state of the global clean and improved cooking sector, ESMAP GACC (2015) 1–179.
[5] A.M. Mobarak, P. Dwivedi, R. Bailis, L. Hildemann, G. Miller, Low demand for nontraditional cookstove technologies, Proc. Natl. Acad. Sci. U. S. A. 109 (27) (2012) 10815–10820.
[6] K.S. Thacker, C.A. Mattson, A global review of End user needs: establishing the need for adaptable cookstoves, Proceedings of the 4th IEEE Global Humanitarian Technology Conference (2014) 649–658.
[7] K.S. Thacker, K.M.C. Barger, C.A. Mattson, Balancing technical and user objectives in the redesign of a peruvian cookstove, Dev. Eng. 2 (2015) 12–19.
[8] N.G. Johnson, Village energy system dynamics of an isolated rural West African village, Digit. Repos. Grad. Theses Diss. Iowa State Univ. 1 (2012) 1–239.
[9] B. Victor, Sustaining culture with sustainable stoves: the role of tradition in providing clean-burning stoves to developing countries, Cons. J. Sustain. Dev. 5 (1) (2011) 71–95.
[10] S.E.T. Muneer, M.E.W. Mukhtar, Adoption of biomass improved cookstoves in a patriarchal society: an example from Sudan, Sci. Total Environ. 307 (1–3) (2003) 259–266.
[11] A.E. Wood, C.A. Mattson, 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, (2016), pp. 1–10.
[12] N.A. MacCarty, K.M. Bryden, An integrated systems model for energy services in rural developing communities, Energy 113 (2016) pp. 536–557.
[13] S. Abdelnour, C. Pemberton-Pigott, For cook and climate: certify cookstoves in their contexts of use, Energy Res. Soc. Sci. 44 (2018) 196–198.
[14] N.D. Moses, N.A. MacCarty, Usability Testing Protocol for Cookstoves, Oregon State University, 2018, https://humanitarian.engineering.oregonstate.edu/project-page/usability-testing-protocol-cookstoves.
[15] H.R. Hartson, T.S.T. Andre, R.R.C. Williges, Criteria for evaluating usability evaluation methods, Int. J. Hum. Interact. 13 (4) (2001) 1–35.
[16] K. Hornbæk, Dogmas in the assessment of usability evaluation methods, Behav. Inf. Technol. 29 (1) (2010) 97–111.
[17] J. Kim, S.H. Han, A methodology for developing a usability index of consumer electronic products, Int. J. Ind. Ergon. 38 (3–4) (2008) 333–345.
[18] D. Wixon, The unfulfilled promise of usability engineering, J. Usability Stud. 6 (4) (2011) 198–203.
[19] H.R. Hartson, J.C. Castillo, Remote evaluation for post-deployment usability improvement, Proceedings of the Working Conference on Advanced Visual Interfaces - AVI' 98, (1998), p. 22.
[20] T. Clemmensen, Q. Shi, J. Kumar, H. Li, X. Sun, P. Yammiyavar, Cultural usability tests–how usability tests are not the same all over the world, LNCS 4559 (2007) 281–290.
[21] D. Sy, Adapting usability investigations for agile user-centered design, J. Usability Stud. 2 (3) (2007) 112–132.
[22] H. Kiewe, How may I help you? An ethnographic view of contact-center HCI, J. Usability Stud. 3 (2) (2008) 74–89.
[23] D.R. Millen, Rapid Ethnography: Time Deepening Strategies for HCI Field Research, (2000).
[24] M.E. Bentley, G.H. Pelto, W.L. Straus, D.A. Schumann, C. Adegbola, E. de la Pena, G.A. Oni, K.H. Brown, S.L. Huffman, Rapid ethnographic assessment: applications in a diarrhea management program, Soc. Sci. Med. 27 (January (1)) (1988) 107–116.
[25] S.C. Scrimshaw, E. Hurtado, Rapid Assessment Procedures for Nutrition and Primary Health Care. Anthropological Approaches to Improving Programme Effectiveness, Los Angeles California University of California Los Angeles Latin American Center 1987, 1987.
[26] S. Pink, J. Morgan, Short-Term Ethnography: Intense Routes to Knowing vol. 36, John Wiley & Sons, Ltd, 2013, pp. 351–361 no. 3.
[27] K.M. Carley, M.W. Bigrigg, B. Diallo, K.M. Carley, M.W. Bigrigg, B. Diallo, B. Diallo, Data-to-model: a mixed initiative approach for rapid ethnographic assessment, Comput. Math. Organ. Theory 18 (2012) 300–327.
[28] H. Knoblauch, Sociological ethnography, strangeness and alterity, Forum Qual. Soc. Res. 6 (3) (2005) Art. 44, p. 14pp, Sep.
[29] E. Isaacs, The value of Rapid ethnography, Advancing Ethnography in Corporate Environments: Challenges and Emerging Opportunities vol. 1, (2012), pp. 92–107.
[30] H. Bernard, Research Methods in Anthropology: Qualitative and Quantitative Approaches, 4th ed., AltaMira Press, Lanham, 2011.
[31] L. Herman, Towards effective usability evaluation in Asia: cross-cultural differences, Proc. Sixth Aust. Conf. Comput. Interact. (1996) 135–136.
[32] J. Loo, L. Hyseni, R. Ouda, S. Koske, R. Nyagol, I. Sadumah, M. Bashin, M. Sage, N. Bruce, T. Pilishvili, D. Stanistreet, User perspectives of characteristics of improved cookstoves from a field evaluation in Western Kenya, Int. J. Environ. Res. Public Health 13 (January (2)) (2016) 167.
[33] A. Jagadish, P. Dwivedi, In the hearth, on the mind: cultural consensus on fuelwood and cookstoves in the middle Himalayas of India, Energy Res. Soc. Sci. 37 (April) (2017) 44–51 2018.
[34] L. Iessa, Y.A. De Vries, C.E. Swinkels, M. Smits, C.A.A. Butijn, What's cooking? Unverified assumptions, overlooking of local needs and pro-solution biases in the solar cooking literature, Energy Res. Soc. Sci. 28 (2017) Elsevier, pp. 98–108, 01-Jun-.
[35] D. Chatti, M. Archer, M. Lennon, M.R. Dove, Exploring the mundane: towards an ethnographic approach to bioenergy, Energy Res. Soc. Sci. 30 (2017) 28–34.
[36] M.J. Herington, P.A. Lant, S. Smart, C. Greig, E. van de Fliert, Defection, recruitment and social change in cooking practices: energy poverty through a social practice lens, Energy Res. Soc. Sci. 34 (2017) 272–280.
[37] G. Davies, Clean energy product markets in sub-Saharan Africa: complex market devices and power asymmetries, Energy Res. Soc. Sci. 42 (2018) 80–89.
[38] E. Adkins, E. Tyler, J. Wang, D. Siriri, V. Modi, Field testing and survey evaluation of household biomass cookstoves in rural sub-Saharan Africa, Energy Sustain. Dev. 14 (2010) 172–185.
[39] M. Jürisoo, F. Lambe, M. Osborne, Beyond buying: the application of service design methodology to understand adoption of clean cookstoves in Kenya and Zambia, Energy Res. Soc. Sci. 39 (May) (2018) 164–176.
[40] Y. Malakar, C. Greig, E. van de Fliert, Resistance in rejecting solid fuels: beyond availability and adoption in the structural dominations of cooking practices in rural India, Energy Res. Soc. Sci. 46 (December) (2018) 225–235.
[41] P.P. Otte, A (new) cultural turn toward solar cooking - evidence from six case studies across India and Burkina Faso, Energy Res. Soc. Sci. 2 (June) (2014) 49–58.
[42] O. Akintan, S. Jewitt, M. Clifford, Culture, tradition, and taboo: understanding the social shaping of fuel choices and cooking practices in Nigeria, Energy Res. Soc. Sci. 40 (June) (2018) 14–22.
[43] Y. Malakar, Studying household decision-making context and cooking fuel transition in rural India, Energy Sustain. Dev. 43 (2018) 68–74.
[44] Y. Wang, R. Bailis, The revolution from the kitchen: social processes of the removal of traditional cookstoves in Himachal Pradesh, India, Energy Sustain. Dev. 27 (2015) 127–136.
[45] I. Miremadi, Y. Saboohi, S. Jacobsson, Assessing the performance of energy innovation systems: towards an established set of indicators, Energy Res. Soc. Sci. 40 (December) (2017) 159–176 2018.
[46] A. Kar, H. Zerriffi, From cookstove acquisition to cooking transition: framing the behavioural aspects of cookstove interventions, Energy Res. Soc. Sci. 42 (2018) 23–33.
[47] N.D. Moses, M.H. Pakravan, N.A. MacCarty, Development of a practical evaluation for cookstove usability, Energy Sustain. Dev. 48 (2019) pp. 154—163.
[48] E. García-Frapolli, A. Schilmann, V.M. Berrueta, H. Riojas-Rodríguez, R.D. Edwards, M. Johnson, A. Guevara-Sanginés, C. Armendariz, O. Masera, Beyond fuelwood savings: valuing the economic benefits of introducing improved biomass cookstoves in the Purépecha region of Mexico, Ecol. Econ. 69 (12) (2010) 2598–2605.
[49] I. Ruiz-Mercado, O. Masera, H. Zamora, K.R. Smith, Adoption and sustained use of

improved cookstoves, Energy Policy 39 (12) (2011) 7557–7566.
[50] T. Urmee, S. Gyamfi, A review of improved Cookstove technologies and programs, Renew. Sustain. Energy Rev. 33 (2014) 625–635.
[51] D. Kimemia, A. Van Niekerk, Cookstove options for safety and health: comparative analysis of technological and usability attributes, Energy Policy 105 (2017) 451–457.
[52] M.P. Kshirsagar, V.R. Kalamkar, A comprehensive review on biomass cookstoves and a systematic approach for modern cookstove design, Renew. Sustain. Energy Rev. 30 (2014).
[53] International Organization for Standardization, ISO 9241:11 Ergonomics of Human System Interaction, (1998).
[54] N.G. Johnson, K.M. Bryden, Field-based safety guidelines for solid fuel household cookstoves in developing countries, Energy Sustain. Dev. 25 (2015) 56–66.
[55] Center for energy development; energy institute at Colorado State University, Cookstove Durability Protocol (CDP), (2014).
[56] International Organization for Standardization, ISO/IWA 11: 2012 Guidelines for Evaluating Cookstove Performance, (2012).
[57] J. Nielsen, Usability inspection methods, Conf. Companion Hum. Factors Comput. Syst. - CHI' 94 25 (1) (1994) 413–414.
[58] C. Garland, M. Johnson, K. Jagoe, T. Ramanathan, Cookstove Field Study Resources Information, Tools, and Sample Forms for Researchers, Communities, and Organizations List of Contributors, (2016).
[59] International Organization for Standardization, ISO/TC 285 Clean Cookstoves and Clean Cooking Solutions, International Organization for Standardization, Geneva, Switzerland, 2018.
[60] R. Bailis, R. Edwards, Kitchen Performance Test (KPT), Shell Foundation, 2007.