ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #36948

Developing design ethnography interviewing competencies for novices

Micki Grover

Michaela "Micki" Grover is a master's student studying product design at the University of Minnesota. She has a Bachelor of Science in Mechanical Engineering and three years of professional experience as a manufacturing engineer. Her research publication topics include microgrid electrical systems, inclusive design, and ethnographic interview methods. She has a special interest in toy design.

Natasha Wright

Assistant Professor, University of Minnesota

Jennifer Margaret Hoody

Jennifer Hoody has experience and research interests centered on sustainable and global development with the goal to pursue a career centered at the intersection of environmental, climate, and social justice from an engineering mindset. Her past research focused on topics such as renewable energy, energy access, water desalination, human-centered design, and medical device development. She gained a passion for this field of work through her undergraduate research investigating energy reduction programs for underserved communities and through her research experience on solar brine management as a participant in a National Science Foundation research experience for undergraduates program. She earned her Bachelor's in Mechanical Engineering with a minor in Sustainability, Energy, and the Environment at the University of Dayton in 2020. In 2022, she received her MS in Civil and Environmental Engineering at the University of California, Berkeley within the Energy, Civil Infrastructure, and Climate program. During her master’s program, she gained further experience conducting research and working with underserved communities on a local and national level. Jennifer will be beginning her PhD at the University of Minnesota, Twin Cities in the fall of 2022, where she plans to continue research focused on engineering for global and sustainable development, with specific interests in the water-energy-food nexus.

Carlye Lauff

Dr. Carlye Lauff is an Assistant Professor of Product Design at the University of Minnesota. She earned her PhD from the Department of Mechanical Engineering at the University of Colorado Boulder, where she was a National Science Foundation Graduate Research Fellow studying the role of prototypes in companies. Carlye’s research is in the field of Design Theory and Methodology, and she studies how designers engage in the product development process and then improves tools and methods to support them.

Developing design ethnography interviewing competencies for novices

1.0 Introduction

Ethnographic research principles are used in design projects to cultivate an empathetic understanding of stakeholders' needs. A common ethnographic research practice is semi-structured, direct-dialogue interviews between design team members and stakeholders of the project. Students who conduct interviews for design research projects are often "novices," with little or no prior experience in ethnographic research. Interviewers who have been trained in interviewing have been shown to produce more accurate and in depth information [1]. This study poses the question: How might we verify that novices have baseline competency for conducting ethnographic design interviews before engaging in interviews for the design project?

Ethnographic research in design differs from other user research methods, such as surveys or focus groups, in that it seeks to understand how people actually think, feel, and act in context as opposed to what they may claim [2]–[4]. In organizations and industry, ethnography is a practice used to drive innovation in design [2]. Design ethnography is a useful tool, and to competently interview using ethnographic principles takes knowledge and experience.

This study recruited novice interviewers for a training program on ethnographic research principles as well as design thinking, empathy mapping, stakeholder mapping, participatory design, service design, and co-design. The training prepared the students for their Research Assistant positions in a design project. The purpose of the project was to develop a potential service model in sub-Saharan Africa to introduce a new dialysis technology. The research project is made up of the following two aims:

Aim 1: Perform design ethnography case studies in sub-Saharan Africa (Uganda and Zimbabwe) that will provide the data needed to inform the development of an economically, culturally, and technically accessible peritoneal dialysis (PD) treatment program and service model.

Aim 2: This study will serve as a pilot study for future work aimed at addressing the following questions: How might we verify that novices have baseline competency for conducting ethnographic design interviews? In what ways do interview and co-creation workshop outcomes vary based on the background of the novice interviewer?

While the students were recruited to participate in Aim 1, this paper is focused on the first research question of Aim 2: How might we verify that novices have baseline competency for conducting ethnographic design interviews? Aim 1 and the second research question of Aim 2 will be addressed in the continued research.

In the study, professionals experienced in ethnographic design interviewing reviewed and scored practice interviews conducted by the novice students during the training program. Those same interviews were evaluated using software to code and calculate scores derived from literature-supported interview metrics. The scores given by the experts were compared to the scores generated by the metrics and software calculations. The comparison of the two scoring

techniques explored the usefulness and limitations of each approach used to verify novice interviewer readiness.

2.0 Background

Design ethnography differs from the ethnographic approaches of anthropologists. Design ethnography uses ethnographic principles, such as employing empathy and generating understanding in the context of the subjects' daily lives, but does so in a less immersive manner and over shorter time periods [5]. In product design research, design ethnography is an approach chosen for generating intercultural understanding [6]. This understanding is then utilized to develop designs that more accurately meet the needs of those studied.

One research method in design ethnography is semi-structured, direct-dialogue interviewing. Such interviews can be used for a variety of purposes, including uncovering latent user needs, developing product specifications, finding new contacts and stakeholders, building a relationship with stakeholders, generating a better understanding of the context of the problem, and exploring the desirability, viability, and feasibility of a product or solution. Often multiple of these purposes are met in a single interview.

Students and other novices typically learn more by doing than through more passive learning activities [7]. For this reason, practice interviews are used during training. Additionally, arranging for low-stakes interviews at the beginning of a study, such as interviews with less prominent stakeholders, is a strategic approach to preparing students for the more high-stakes subjects. In some instances, it can be helpful for a novice interviewer to be paired with a more experienced interviewer to assist in their learning during field work. As the students gain more interview exposure, they strengthen their interviewing skills.

There are many documented best practices in interviewing. Soft skills such as generating good rapport during the interview and including more casual conversation to help the interviewee relax can increase the likelihood of uncovering helpful information [3], [8]. Other soft skills for interviewing include building trust, showing compassion and care, and using positive facial expressions and gestures to encourage the interviewee. Technical skills, such as asking open ended questions and repeating back what is heard, give permission to the interviewee to guide the interview towards previously unanticipated and potentially crucial insights [2]. Other technical skills for interviewing include being a good listener, using reflective and active listening techniques, asking follow-up questions to dig deeper into topics, and probing for stories and emotions. Using the recommendations for successful interviews documented in literature, criteria for a quality interview can be taught to novice interviewers.

3.0 Research Design and Approach

3.1 Participants

Students from the University of Minnesota (UMN) and from universities in sub-Saharan Africa (SSA) were recruited to participate in the study as Research Assistants with a digital flier and an online interest form. The University of Minnesota students were recruited from the student

groups Engineering World Health and Engineering Without Borders and from the Department of Mechanical Engineering. Students in sub-Saharan Africa were recruited through the Engineering for Change (E4C) Opportunities Portal and E4C fellows alumni Slack channel, as well as through local network university connections. The interest form received fourteen submissions from students at the University of Minnesota and ninety-six submissions from students in sub-Saharan Africa.

Six students from the University of Minnesota and six students from universities in sub-Saharan Africa were chosen through a two-stage process. The first stage was an interview conducted by a member of our research team with the focus of asking follow-up questions based on the student's interest form responses and inquiring about experiences pertinent to the study. The second stage was an informational meeting led by the Research Assistant and a Primary Investigator Research Professor on the study. The purpose of the meeting was to introduce the two aims of the study and answer any additional follow up questions.

The students were selected based on their location, their status as a "novice" interviewer, with novice defined as someone with little to no interview experience, and their willingness to be a part of Aim 1 of the study. The university locations and majors of the selected students are detailed in Table I.

TABLE I: Location and Major of Selected Students

Location	Major
Harare, Zimbabwe	Electronic Engineering
	Electronic Engineering
Kampala, Uganda	Cytotechnology
	Biomedical Engineering
Mbarara, Uganda	Biomedical Engineering
	Medicine and Surgery
Minnesota, United States	Biomedical Engineering
	Biomedical Engineering
	Mechanical Engineering
	Bioproducts and Biosystems Engineering
	Genetics, Cell Biology and Development Minor: Microbiology
	Neuroscience and Economics

3.2 Training Overview

To prepare students for this design project, the research team developed an online Design Ethnography Training Program, which consisted of five virtual workshops, assignments based on each workshop, and feedback provided after each assignment. Each workshop session occurred over Zoom utilizing online collaborative tools, like Lucidchart, and lasted between 1-2 hours. The goal of the training program was to prepare the participants to conduct effective ethnographic interviews and to co-create stakeholder maps with their interviewees. These skills were developed to be used to meet the Aim 1 research objectives. A breakdown of the training topics is shown in Table II.

TABLE II: Design Ethnography Training Program Workshop Topics

Workshop 1	Introductions, Design Thinking
Workshop 2	Interviewing
Workshop 3	Interviewing (continued), Stakeholder Mapping
Workshop 4	Service Models, Field Interviews
Workshop 5	Participatory Design Sessions

After Workshop 2, the students were instructed to conduct roughly 10 minute long practice interviews in pairs. For these initial practice interviews, students were paired by their home country for ease of organizing times for interviews. For each interview, one student was the interviewer while the other was the note taker. The note taker was also the recruiter of the interviewee for that interview. The twelve students took the role of the interviewer twice each for a total of twenty four interviews. Later in the training, the students conducted 30 minute practice interviews. For the 30 minute practice interviews, students were paired based on a mix of both students being from the same country and then also across countries. This aided in the second part of research Aim 2, which is related to how background and cultural differences influence interviews. This will be discussed in future research.

The prompt for the practice interviews was as follows: *"You recently began a new job at a popular food company that produces and sells pre-packaged meals. They have begun a new initiative in an effort to increase their products offerings and provide healthier options. To ensure their new products will be well-received by customers, your team is working to gain information about the current state of healthy food options. Your team is also tasked with understanding what drives customers' grocery shopping decisions, such as what current limitations and barriers exist that prevent individuals from choosing healthier alternatives, what access to healthy options currently look like, and what the ideal options would be to make healthier purchases. Before moving forward in the product design process, your manager has assigned you the task of conducting qualitative interviews with community members to help answer these questions and gain insights that will inform the direction of product development."*

All students were instructed to read the IRB-approved verbal consent script and receive documented verbal consent before proceeding with the interview. The verbal consent process was a new concept for the students and therefore was not perfectly executed in all instances. Nine of the twenty four interviews did not receive properly documented verbal consent and as a result could not be used in the study.

3.3 Interview Metrics

This pilot study is a starting place for establishing what metrics might validate certain interviewing skills. The metrics in this study (see Table III) were formed based on best practices from literature and interviewing skills identified by the experience of the research team. The metrics were iterated upon with the research team and outside experts until they were refined into a list that felt holistic and complete.

In addition to best practices and interviewing skills for an interview, worst practices were also identified. These worst practices were known as “detractors” from the interview, which means that in some way they detract from the quality of the interview. The detractors identified by the research team are: expressing extreme emotion, asking a leading question, using signposting unnecessarily, interrupting the interviewee, showing negativity, revealing an opinion, reacting defensively, and showing bias. Alone, a couple of the detractors might not drastically impact the outcome of an interview. However, if multiple detractors occur, then it can negatively impact the interview quality. For example, leading questions, embedded bias, or opinions given by the interviewer can unduly influence the answers supplied by the interviewee. Keeping detractors to a minimum is an indicator of an interviewer’s skill.

3.4 Data Collection - Novice Practice Interviews

All twenty four practice interviews were conducted over Zoom with automatic recording and transcription options turned on. Zoom’s transcription program provides timestamps whenever there is a pause or change of speaker. Transcription dialogue errors were corrected by a member of the research team while reviewing the recording. The interview videos and transcriptions were imported into NVivo (QSR International), a qualitative data analysis software. The transcriptions were paired with their corresponding videos so that they could be reviewed simultaneously in the same user interface.

3.5 Data Analysis - NVivo Coding of Interviews

NVivo has tools to “code” sections of the transcript. Coding is the process of labeling sections of the transcript in an attempt to elucidate themes and relationships between them. The labels were applied by a research assistant and were based on the interview skills listed in the interview metrics created (Table III). For example, if the interviewer asked an open-ended question, the text of the question was selected in NVivo and the code “Open-Ended Questioning” was applied. Once the entire interview was coded, NVivo used the timestamps of the coded interview sections to calculate the “percent coverage” of each code. The percent coverage of a code is the duration of the sum of that code over the total duration of the interview. For example, if the percent

coverage of the code “Talking” is 20%, then the interviewer was talking for 20% of the interview.

Every interview had time before the interview questions began and after the interview questions ended that was removed for duration calculations of coverage data in the body of the interview. For consistency, the duration of the body of the interview was considered to begin after the consent process was complete and upon the first interview question being asked and to end after the interviewee’s answer to the last interview question. To assist in analyzing the interviews, the research team divided the interviews into three components: opening, body, and closing. The team observed that the opening and closing occur only a small fraction of the time when compared to the overall interview. The body of the interview was the majority of the time, since that is when most of the interview questions are asked.

The body of the interview was coded for Active Listening Vocalizations, Follow-Up Questions, Reflective Listening, Open-Ended Questioning, Encouraging Storytelling, and Helpful Signposting codes to determine if the interview met the corresponding metrics. Some criteria for the body of the interview were met by the existence of at least one code. These were considered “Yes or No” metrics. The body of the interview also included metrics related to the percent coverage of certain codes. For example, the metric for Open-Ended Questions requires greater than 70% of the time spent questioning to be spent on asking open-ended questions. This was calculated by taking the time spent asking open-ended questions over the total time spent asking any questions (open-ended and closed-ended).

The opening and closing portions of the interview were made up entirely of “Yes or No” metrics. The opening of the interview was coded for Building Rapport, Initial Non-Interview Questioning, and Stage Setting. The closing portion of the interview was coded for Closing Remarks, Asking Partner for Questions, and Asking Interviewee for Questions. If the elements of the opening and closing portion of the interview were present, the criteria was met. An interview metric that was removed from the practice interviews but would be evaluated in field interviews was “Ask for New Contacts.”

After the interviews were coded, the coding results were documented in a spreadsheet. The spreadsheet detailed which metrics were met or not met per interview. The data from the spreadsheet is located in the Appendix (Tables V-X).

The calculation for each interview score was generated using the metrics derived from literature (Table III). The various interview criteria were weighted to reflect their importance to the overall success of the interview. The body of the interview is the largest portion of the interview and where interviewee insights are requested and received. For this reason, the research team decided to weight the criteria based on where the criteria occurred (opening, body, closing). The weight was chosen based on the time spent on each portion of the interview: 70% for the body of the interview and 30% for the combined opening and closing of the interview. These values are under consideration for adjustment after more interview data has been gathered and evaluated.

The opening portion of the interview had three metrics, the body had six metrics, and the closing had three metrics. The individual criteria which made up the opening, body, and closing of the

interview were each weighted to reflect their significance to the interview (Table III). The weighting of the individual criteria was based on the combined experience of the research team. After reviewing all the interviews, the experts reflected on their reviews to rank the priority of the individual criteria. Based on their rank, each criteria was given a weight between one and three points. For example, use of "Helpful Signposting" would earn the interview one point while "Reflective Listening" would earn three points. These weights will be reevaluated based on the results of future interviews within the study.

To respect the diversity of approaches interviewers must take depending on who they are interviewing, the weighting of the criteria allowed for more than one path to a perfect score. Both the body and the combined opening/closing of the interview had 12 points possible when 10 points constituted a perfect score for each interview section. If the interview received over 10 points in either the body or the combined opening/closing of the interview, the score was reduced back down to 10 so that no interview could earn over 100%. The points from the body of the interview were then multiplied by 0.7, the points from the opening/closing of the interview were multiplied by 0.3, and the two were summed for the total.

The total was reduced if there were any interview detractors, like a leading or biased question being asked. For each detractor present in the interview, 3% (0.3 points) was subtracted from the total score. The chosen value of 3% is based on the impression that three detractions would equal a reduction of 9%, which would translate to about 1 point out of a 10 point final rating being removed. Further study will determine if a constant 3% per detraction is appropriate. One option proposed by the research team was for the first 3 detractors to be forgiven. This would account for potential coder misinterpretation and sometimes necessary or unavoidable interruptions.

TABLE III. Interview Metrics, Values, and Weights

Interviewing Skill	**Interview Section**	**Metric**	**Value**	**Weight per Section**
Building Rapport	Opening	Part of opening used to build rapport	Yes	2
Initial Non-Interview Questioning	Opening	Non-interview questions asked	Yes	2
Stage Setting	Opening	Opening statements used to set the stage for the interview	Yes	3
Active Listening Vocalizations	Body	Active listening vocalizations used (Ex. "Uh-huh", "I see", "Okay")	Yes	1
Follow-Up Questions	Body	Percent of time talking spent asking follow-up questions	>50%	3

Reflective Listening	Body	Reflective listening demonstrated by repeating back one or more answers	Yes	3
Open-Ended Questioning	Body	Percent of time time spent on open ended questions versus closed ended	>70%	2
Encouraging Storytelling	Body	Percent of time spent talking used to encourage story telling	>30%	2
Helpful Signposting	Body	Helpful signpost used (Ex. signposting used for transitions or clarification)	Yes	1
Closing Remarks	Closing	Closing remarks used to conclude the interview	Yes	3
Asking Partner for Questions	Closing	Interviewer asks if their partner has any questions	Yes	1
Asking Interviewee for Questions	Closing	Interview asks interviewee for any questions	Yes	1
Extreme Emotion	Detractor	Extreme emotion expressed by interviewer	No	1
Leading Question	Detractor	Leading question asked	No	1
Unhelpful Signpost	Detractor	Unhelpful signpost used (Ex. signposting used that is unnecessary or survey-like)	No	1
Interruption	Detractor	Interruption of interviewee	No	1
Negativity	Detractor	Negativity expressed by interviewer	No	1
Opinion	Detractor	Opinion given by interviewer	No	1
Defensiveness	Detractor	Defensiveness expressed by interviewer	No	1

3.6 Expert Evaluations

The validity of the interview rating based on the NVivo coding was determined by comparing it to the ratings given by interview experts. The interview experts in this study were the two research professors leading the study and one lead researcher developing the workshop materials. Combined, the three experts have more than twenty years of experience conducting interviews for design and research projects. The interview experts have a mixed background in mechanical engineering, product design, and aspects from the social sciences like anthropology and psychology.

The three interview experts divided the interviews between them and evaluated and rated their assigned interviews. The three experts used an open-ended form based on the interviewing skills to give feedback to the students about each aspect of the interview. This form was discussed before evaluations began to ensure all experts understood the categories. At the end of the evaluation form, there was a category to numerically evaluate the students' interview on a scale of one to ten, with one meaning the interviewer conducted a poor interview and 10 meaning the interviewer conducted an excellent interview. The experts associated a score of one to be more like a "survey". A survey is a standalone document that does not allow for any improvisation. If a student conducted an interview like a survey, this meant the student asked bulleted questions without leaning into their active and reflective listening that was taught in the previous training session. A score of a 10 fulfilled all of the interview requirements taught during the training. It should be noted that no students received a 10 and that they fell somewhere between 1 and 7, as is shown in the comparison (Table V).

Later in the research project, two of the experts reflected on their process for scoring in a written memo. Both experts placed the greatest emphasis on the body of the interview, which is the question-asking portion of the interview after opening statements and before the closing remarks. The experts were both looking for examples of active and reflective listening in particular, skills that set an interview apart from a survey. When a student simply read their questions off of a prepared script in order, they did not demonstrate active or reflective listening and were therefore given a lower score.

One of the experts placed a mid-tier level of importance on the opening of the interview and very little importance on the closing while the other expert did the opposite. The expert who placed greater importance on the opening was taking into special consideration the effect that setting the stage of the interview and building rapport have on the rest of the interview. The expert who placed greater importance on the closing was thinking about how vital it is to refer to one's interview partner for additional questions or general acknowledgement, to thank the interviewee, and to ask for additional contacts or a future follow-up interview. When discussing their rankings, both experts recognized the importance of the opening and closing stages of the interview and agreed on considering those interview aspects equal for the software scoring portion of the study.

The experts both considered potential interview mistakes, such as leading questions and not being neutral, to have the least effect on their score. If the mistakes showed up in an obvious or repeated way they would have a negative impact on the score, however this did not occur often,

partially due to the short length of the practice interviews. The experts did not give points for the mistakes not being present but would potentially remove points if these mistakes ("detractors") occurred multiple times. This is reflected in the coded scoring by giving those criteria a zero if not present and a negative score of -3% per occurrence.

The difference between the coded score and the expert score was calculated by subtracting the coded rating from the expert rating. Therefore, if the rating was positive, that meant the expert rating was higher than the coded rating. If the rating was negative, that meant the expert rating was lower than the coded rating. The desired output was zero as this would mean the method of calculating the interview rating was perfectly aligned with expert opinion.

4.0 Results and Discussion

4.1 Expert Scores

The students were determined to be ready for field interviews if they achieved an average interview score of 7 from the interview experts. The 10 minute interviews analyzed in this study were conducted after Workshop 2, the students' first independently conducted interviews during the training. Of the seven students evaluated, two students scored a 7 while five students scored below 7. These scores represented the skill level of the novice interviewers early on in the training. Following the 10 minute interviews, each student received personalized instructor feedback before continuing on with the remainder of the training. Later in the training, the students conducted another round of practice interviews, each roughly 30 minutes in length, which were also evaluated to determine readiness. The data from the second round of practice interviews is not included in this paper.

4.2 Coded Scores vs. Expert Scores

The interview score generated by the coding and calculating process was compared to the score given by the interview experts. These scores were averaged from the two interviews per student, except for in the case of student number 3. Student number 3 only had one interview that could be used due to improperly requested consent during the other interview. The chart in Figure 1 shows the agreement and error between the coded and expert scores.

Points above the diagonal line had a greater coded score than an expert score. There are two data points above the line, meaning the experts gave a worse score compared to the software coded score. Data points below the diagonal line had the opposite result; they had a lower coded score than the expert score. There are four data points below the line, meaning that experts gave a better score when compared to the software coded score. One data point fell exactly on the line, meaning the score was the same between the experts and the software coded score.

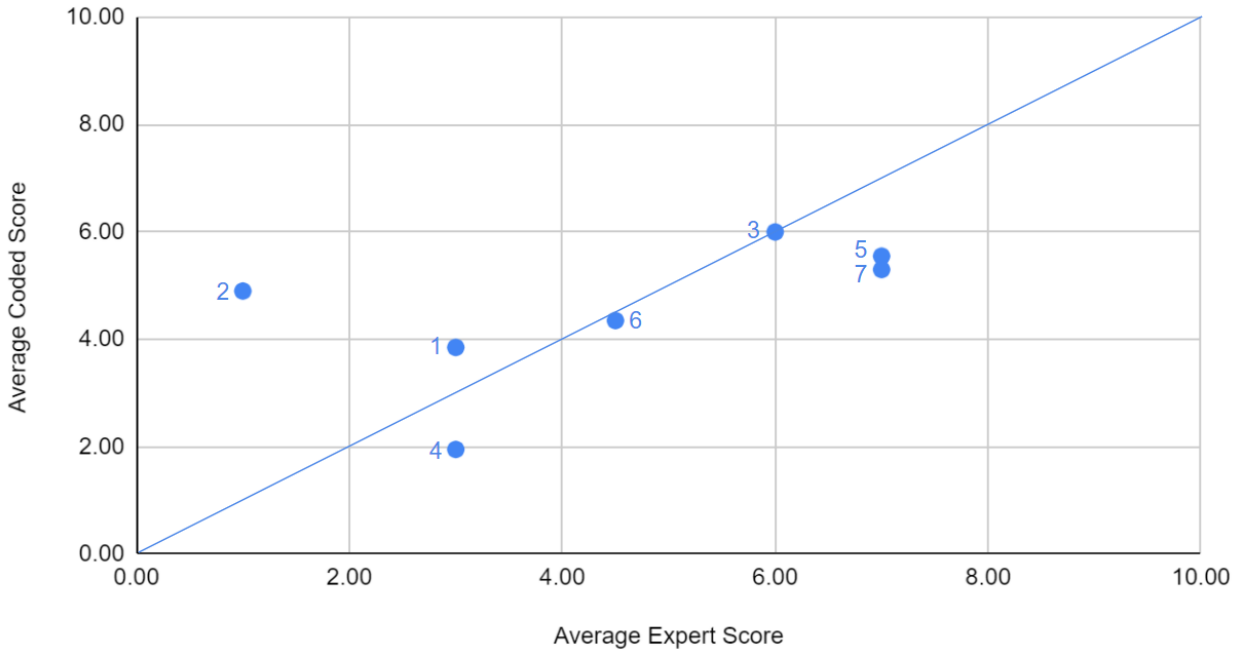

Figure 1. Average Coded Score vs. Average Expert Scores per Student (numbered 1-7)

There were differences between most of the expert reviewed interview scores and the software coded scores. Potential rationale for these differences was explored. Further exploration of the methods used to determine interviewer readiness will be included in the later stages of the design research project.

Six of the seven interviews that had higher coded scores than expert scores met the criteria for Open-Ended Questioning (2 points multiplied by 0.7 for being in the interview body). The Open-Ended Questioning criteria was met when the interviewer used over 70% of the time spent asking questions on asking open-ended questions as opposed to closed-ended questions. Before conducting the practice interviews, the students in the training participated in a classroom brainstorming session to generate a list of mostly open-ended interview questions to use. If the students spent the majority of their interview on the prepared questions, they likely met the Open-Ended Questioning metric. However, the experts recognized when the prepared questions were used and were not as inclined to award credit for their use due to the interview feeling more like a survey asking bulleted pre-prepared questions than leaning into their active and reflective listening training. In the future, the research team is considering having the individual teams prepare for their interviews independently to better evaluate their choices of questions.

Three of the four interviews that received an expert score over 1 point higher than their coded score did not meet the criteria for Open-Ended Questioning. This may reflect that students who used less of the prepared questions from the classroom discussion also used less Open-Ended Questions. Using less of the prepared questions may have been rewarded by the experts while potentially causing those students to not meet the Open-Ended Questioning criteria. If the criteria for Open-Ended Questioning was lowered from 70% to 60%, all four of the interviews would have been awarded the points in the coded score. This may indicate a needed change in the coded

scoring criteria and/or the effect of the experts knowing which questions were scripted for the interviews.

The two students with higher coded scores than expert scores conducted the shortest interviews on average. The duration of the interview body, which begins when the first interview question is asked and ends after the last interview question is answered, was about 5 minutes on average for Student 1 and 2.7 minutes on average for Student 2. These short interviews may not have been regarded as highly by the experts due to missed opportunities to demonstrate interviewing skills such as using active and reflective listening to dig deeper or probing for stories, which often lengthened total interview time. The method of using coding to score interviews cannot recognize missed opportunities, only the codes that exist.

The four interviews that received an expert score over 1 point higher than their coded score had higher coded scores for the opening/closing portions of their interviews but were weighed down by the score for the body portion of their interview (Table IV). This may indicate that the experts put more weight on the opening/closing portions of the interview than the coded scoring formula. The body portion of the interview makes up 70% of the coded score and the opening/closing portions of the interview only make up 30%. The formula may need to be adjusted. Three of these interviews were also affected by detractors. The coded scores for interviews number 13 and 14 were reduced by nearly an entire point due to having three detractors each. If the negative value for each detractor was reduced, or if the first couple detractors were forgiven, the coded scores would have been closer to the expert scores in these cases.

TABLE IV: Evaluation of Interviews with an Expert Score >1 point than their Coded Score

Interview #	Student #	Body Score (Out of 10)	Opening / Closing Score (Out of 10)	Detractors (-0.3 each)	Coded Score	Expert Score
6	4	0	3	1	0.6	3
9	5	3	9	0	5	7
13	7	6	8	3	5.7	7
14	7	4	10	3	4.9	7

Student 3 was the only student who had the exact same score from the experts as was generated by the coded scoring method and was also the only student with only one interview available for evaluation due to an improper interview consent process with their other interview. The lack of averaging for student number 3's final score may have contributed to their result being exactly 6 points for both methods.

Though student number 6 did not have an exact match between the expert's score and the coded score, the difference was only 0.15 points. The large number of interview criteria divides the coded score into more detailed percentages, which makes exact matches unlikely. One area for future consideration would be to determine what amount of variation is small enough to be considered a match between the expert score and the coded score. The coded score could also be

rounded up or down to the nearest whole number to match the significant figures of the expert scores.

While students were asked to complete 10 minute practice interviews, many of them were under that required time (Appendix Table X). This lack of achieving a full 10-minute interview is an interesting observation, and could indicate potential factors like lack of skills and confidence for the novice in conducting the interview or disconnect from the practice scenario to engage in a meaningful interview. This is a topic that will be considered in the future research. Additionally, 10 minutes is a very limited time to demonstrate the full range of interviewing skills taught during the training. It also reduces the amount of opportunities for mistakes, which may inaccurately represent the readiness of the student to conduct longer interviews. The shortness of the interviews also impacts the NVivo coverage calculations of the codes. If the entire interview is very short, each sentence holds more weight for duration-based criteria. In general, longer interviews would be a more accurate portrayal of the readiness of the students to conduct ethnographic design interviews.

The greatest outlier was student number 2 (see Figure 1). Student 2 received an average coded score of 5 but an expert score of 1. This could potentially be explained by the duration of the interviews performed by student number 2, which were less than all other students. The duration of the body of the interviews were 1.94 and 3.43 minutes. Though the short interview afforded very little information to be gathered from the interviewee, the interviewer managed to pass about half of the interview criteria based on the codes. This may indicate that the coded scoring is not accurate for very short interviews.

5.0 Limitations

There are limitations to consider in the process of scoring an interview using software. For one, the transcripts generated in Zoom often required editing. This opened up the possibility of human error in addition to software error. Also, the process of coding the interviews was subjective for certain criteria. For instance, there could be debate over if a specific question was “biased” as this was based on the person coding the interview’s interpretation. Finally, many of the scores were calculated based on the amount of time the interviewer was talking in the body of the interview. Technical difficulties using Zoom could require the interviewer to repeat themselves more often than normal, causing an artificial increase in talking time. Additionally, very short interviews make coverage requirements harder to achieve. For example, in the 10 minute practice interviews, it is more difficult for the interviewer to ask enough questions to meet the criteria of having greater than 30% of the talking coverage be asking questions that encourage story-telling. Software-based scoring is useful for evaluating large amounts of data in a structured fashion, however expert evaluation is more equipped to handle ambiguity and nuance.

The experts who evaluated the interviews, though not subject to technical-based limitations, had other potential biases. The three experts in this study divided up the interviews evenly to score. Though the experts put the most emphasis on the body of the interview, one expert put more weight on the opening statements while another put more weight on the closing remarks. This could potentially have led to variation in the expert scores. The experts also knew what the scripted questions were beforehand and therefore knew that use of these questions did not

demonstrate active or reflective listening. The coding process, however, did not make this distinction. In general, deciding whether an interview deserved a 2 vs a 3 or an 8 vs a 9 is more subjective when it is not based on coded calculations, as coded calculations can give quantitative reasoning behind what increases or decreases a score so minutely.

Nine out of twenty four practice interviews could not be used because students did not read the verbal consent script when requesting consent to interview. Afterwards, the research team clarified the verbal consent requirements to ensure proper procedure for future interviews. It is extremely important to ensure the novice interviewers fully understand the official process of requesting and receiving consent. This study revealed that students may have difficulty fully grasping that point in particular.

6.0 Conclusions

In preparation for performing field interviews for a qualitative research study, twelve students participated in an ethnographic research training program. Interview experts rated the practice interviews to determine novice interviewer readiness. To reduce reliance on expert review, a process was created to code sections of the interviews and calculate their ratings based on metrics and criteria weights determined through research. The results of the ratings show general agreement between the coded score and the expert score, but further data analysis is required to refine the scoring process.

The method presented in this paper for measuring the readiness of an interviewer can be used to improve interviewer training and interview evaluation. The method produces comparable scores which can show skill progression, identify strengths and weaknesses, and verify baseline ethnographic interview competency before field interviews are conducted. This makes the method valuable to engineering educators and other qualitative researchers who are working with novice interviewers.

This study is part of an ongoing research project which has advanced to a second stage of practice interviews and field interviews. The scoring method will be refined through the same process presented in this paper of comparing coded scores to expert scores. This evaluation was based on one round of interviews which were all intended to be 10 minutes in length. The second stage of practice interviews were intended to be 30 minutes long and the field interviews were intended to be about 1 hour long. The difference in interview length will present a new variable for consideration on how it may affect scoring. Additionally, there will be a comparison between practice interview scores and field interview scores to determine how well the skills demonstrated during training are utilized in practice.

Data from the field interviews will be used to explore the second question of Aim 2: In what ways do interview and co-creation workshop outcomes vary based on the background of the novice interviewer? The novice interviewers were intentionally paired to draw comparisons during the field interviews. The results of such comparisons may be used to recommend specific configurations of backgrounds of novice interviewer pairs in studies that employ ethnographic interviewing practices, particularly for cross-cultural design projects.

This paper addresses the first question of Aim 2: How might we verify that novices have baseline competency for conducting ethnographic design interviews before engaging in interviews for the design project? This study establishes a baseline evaluation of the core competencies of an ethnographic interview. In practice, the metrics generated and explored by this study can be used to help determine the readiness for novice interviewers, such as students, to conduct ethnographic interviews for design research projects. Additionally, this work will help to answer the second question of Aim 2, which focuses on how interviewee demographics affect interview outcomes, and Aim 1, which focuses on developing a PD service model in sub-Saharan Africa.

References

[1] "CCSG." https://ccsg.isr.umich.edu/chapters/interviewer-recruitment-selection-and-training/ (accessed Jan. 04, 2022).

[2] "101 Design Methods: A Structured Approach for Driving Innovation in Your Organization | Wiley," *Wiley.com*. https://www.wiley.com/en-us/101+Design+Methods%3A+A+Structured+Approach+for+Driving+Innovation+in+Your+Organization-p-9781118083468 (accessed Jan. 04, 2022).

[3] J. Blomberg, M. Burrell, and G. Guest, "An Ethnographic Approach to Design," Lawrence Erlbaum Associates, Inc., 2002. Accessed: Jan. 04, 2022. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:bth-9738

[4] C. Wasson, "Ethnography in the Field of Design," *Hum. Organ.*, vol. 59, no. 4, pp. 377–388, 2000.

[5] B. M. Hanington and B. Martin, *Universal methods of design: 125 ways to research complex problems, develop innovative ideas, and design effective solutions*. S.l.: Rockport Publishers, 2019.

[6] "Quantifying the effects of various factors on the utility of design ethnography in the developing world," *springerprofessional.de*. https://www.springerprofessional.de/en/quantifying-the-effects-of-various-factors-on-the-utility-of-des/16374018 (accessed Jan. 04, 2022).

[7] L. Deslauriers, L. S. McCarty, K. Miller, K. Callaghan, and G. Kestin, "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom," *Proc. Natl. Acad. Sci.*, vol. 116, no. 39, pp. 19251–19257, Sep. 2019, doi: 10.1073/pnas.1821936116.

[8] J. P. Spradley, *The ethnographic interview*. Long Grove, Illinois: Waveland Press, Inc, 2016.

Appendix

TABLE V. Coded Score Calculations and Comparison to Expert Score

Video	Student	Body	Open / Close	Detractors	Weighted Body	Weighted Open / Close	Total	Minus (Detractors)	Coded Score	Expert Score	Difference*
#	#	100%	100%	0.00	70%	30%	100%	0%	10	10	0
1	1	40%	40%	3.00	28%	12%	40%	15%	2.5	3	0.5
4	1	70%	30%	2.00	49%	9%	58%	6%	5.2	3	-2.2
2	2	40%	30%	1.00	28%	9%	37%	3%	3.4	1	-2.4
3	2	70%	60%	1.00	49%	18%	67%	3%	6.4	1	-5.4
7	3	60%	70%	1.00	42%	21%	63%	3%	6.0	6	0.0
6	4	0%	30%	1.00	0%	9%	9%	3%	0.6	3	2.4
8	4	30%	40%	0.00	21%	12%	33%	0%	3.3	3	-0.3
9	5	30%	90%	0.00	21%	27%	48%	0%	4.8	7	2.2
11	5	60%	70%	0.00	42%	21%	63%	0%	6.3	7	0.7
10	6	40%	30%	1.00	28%	9%	37%	3%	3.4	4	0.6
12	6	50%	70%	1.00	35%	21%	56%	3%	5.3	5	-0.3
13	7	60%	80%	3.00	42%	24%	66%	9%	5.7	7	1.3
14	7	40%	100%	3.00	28%	30%	58%	9%	4.9	7	2.1

** The Difference is calculated by subtracting the Coded Score from the Expert Score. Therefore, if the Difference is positive, the Expert Score is greater than the Coded Score.*

TABLE VI. Average Coded Score vs Average Expert Score Comparison

Student	Average Coded Score	Average Expert Score	Average Difference*
1	3.85	3.00	-0.85

2	4.90	1.00	-3.90
3	6.00	6.00	0.00
4	1.95	3.00	1.05
5	5.55	7.00	1.45
6	4.35	4.50	0.15
7	5.30	7.00	1.70

** The Average Difference is calculated by subtracting the Average Coded Score from the Average Expert Score. Therefore, if the Difference is positive, the Average Expert Score is greater than the Average Coded Score.*

Table VII. Interview Body Scores

Video	Student	Active Listening Vocalizations	Score	% Follow Up Questions	Score	Reflective	Score	% Open Ended	Score	% Time Spent Inciting Stories	Score	Helpful Signpost (transition, clarification)	Score
#	**#**	**Y**	**1**	**>50%**	**3**	**Y**	**3**	**>70%**	**2**	**>30%**	**2**	**Y**	**1**
1	1	Y	1	4.82%	0	N	0	76.88%	2	0	0	Y	1
4	1	Y	1	6.63%	0	Y	3	92.79%	2	0	0	Y	1
2	2	Y	1	0.00%	0	N	0	100.00%	2	0	0	Y	1
3	2	Y	1	10.90%	0	Y	3	93.16%	2	0	0	Y	1
7	3	N	0	54.15%	3	Y	3	56.83%	0	4%	0	N	0
6	4	N	0	32.43%	0	N	0	67.10%	0	0	0	N	0
8	4	N	0	22.53%	0	N	0	84.48%	2	0	0	Y	1
9	5	N	0	12.16%	0	Y	3	61.16%	0	7%	0	N	0
11	5	N	0	54.11%	3	Y	3	36.33%	0	0	0	N	0
10	6	Y	1	0.00%	0	N	0	91.83%	2	0	0	Y	1
12	6	Y	1	32.18%	0	Y	3	63.60%	0	0	0	Y	1

13	7	Y	1	32.81%	0	Y	3	88.51%	2	0	0	N	0
14	7	Y	1	7.82%	0	Y	3	61.96%	0	0	0	N	0

TABLE VIII. Interview Opening/Closing Scores

Video	Student	Building Rapport	Score	Initial Questions (non-interview)	Score	Stage Setting	Score	Closing Remarks	Score	Asked Interviewee for Questions	Score	Asked Partner for Questions	Score
#	**#**	**Y**	**2**	**Y**	**2**	**Y**	**3**	**Y**	**3**	**#**	**1**	**#**	**1**
1	1	N	0	N	0	N	0	Y	3	Y	1	N	0
4	1	N	0	N	0	N	0	Y	3	N	0	N	0
2	2	N	0	N	0	N	0	Y	3	N	0	N	0
3	2	N	0	N	0	Y	3	Y	3	N	0	N	0
7	3	Y	2	N	0	N	0	Y	3	Y	1	Y	1
6	4	Y	2	N	0	N	0	N	0	N	0	Y	1
8	4	N	0	N	0	N	0	Y	3	N	0	Y	1
9	5	Y	2	N	0	Y	3	Y	3	N	0	Y	1
11	5	N	0	N	0	Y	3	Y	3	N	0	Y	1
10	6	N	0	N	0	N	0	Y	3	N	0	N	0
12	6	N	0	N	0	Y	3	Y	3	N	0	Y	1
13	7	Y	2	N	0	Y	3	Y	3	N	0	N	0
14	7	Y	2	Y	2	Y	3	Y	3	N	0	N	0

TABLE IX. Interview Detraction Quantities

Video	Student	Extreme Emotion	Leading	Unhelpful Signpost (unnecessary, survey-like)	Interrupt	Negative	Giving Opinion	Defensive	Bias
#	#	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
1	1	0	0	3	0	0	0	0	0
4	1	0	0	1	0	0	1	0	0
2	2	0	0	1	0	0	0	0	0
3	2	0	0	1	0	0	0	0	0
7	3	0	0	0	0	0	0	0	1
6	4	1	0	0	0	0	0	0	0
8	4	0	0	0	0	0	0	0	0
9	5	0	0	0	0	0	0	0	0
11	5	0	0	0	0	0	0	0	0
10	6	0	1	0	0	0	0	0	0
12	6	0	1	0	0	0	0	0	0
13	7	1	0	0	1	1	0	0	0
14	7	0	1	0	1	0	1	0	0

TABLE X. Video and Interview Length

Video	Student	Full Video Length*	Interview Body Length
#	#	minutes	minutes
1	1	17.12	4.88
4	1	14.38	5.24
2	2	7.63	1.94

3	2	9.22	3.43
7	3	15.00	9.47
6	4	14.75	6.58
8	4	10.62	4.44
9	5	14.47	7.03
11	5	20.53	12.77
10	6	9.88	5.26
12	6	20.40	11.69
13	7	11.03	5.45
14	7	14.62	6.43

**The Full Video Length is the raw video length with no edits. This begins when the first attendee joins the Zoom meeting and ends when the last attendee leaves.*