The Development of a New Measure for Community Digital Capacity

TAWANNA R. DILLAHUNT, University of Michigan, School of Information, USA
KERBY SHEDDEN, University of Michigan, Department of Statistics, USA
MILA EKATERINA FILIPOF, University of Michigan, School of Information, USA
SOYOUNG LEE, University of Michigan, School of Information, USA
MUSTAFA NASEEM, University of Michigan, School of Information, USA
KENTARO TOYAMA, University of Michigan, School of Information, USA
JULIE HUI, University of Michigan, School of Information, USA

This article theorizes and proposes a novel construct, *community digital capacity*, to measure collective digital capacity at a community level. Community digital capacity is the extent to which the culture, infrastructure, and digital competence of family and community enable and support digital practices. We address a critical gap in individual digital literacy assessments and address limitations with existing theories that do not show digital inequities in the context of underlying systemic and structural challenges posed by one's social position. Building on insights from Computer Supportive Cooperative Work and Social Computing and Human-Computer Interaction for Development communities, we recognize that digital training initiatives must shift toward critical cultural and social practices that encourage full participation in community affairs. Accordingly, we created 28 items covering three domains—individual, social, and infrastructure. We conducted cognitive interviews with a public housing community to refine the items and capture the construct fully. We assessed their factor structure in two Southeastern Michigan cohorts. After dropping eight items based on contribution to Standardized Root Mean Square Residual (SRMR), the public housing residents exhibit a two-factor structure (SRMR=0.09) consisting of nearly independent factors for the individual and social domains, with all items loading positively on their respective domain. We contribute an initial measure for researchers and practitioners to assess community members' access to shared digital resources and support, offering a tool to assess broader social and structural factors contributing to the digital divide.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: Digital literacy, Measure, Community digital capacity, Social, Public Housing, Refugees

ACM Reference Format:
Tawanna R. Dillahunt, Kerby Shedden, Mila Ekaterina Filipof, Soyoung Lee, Mustafa Naseem, Kentaro Toyama, and Julie Hui. 2025. The Development of a New Measure for Community Digital Capacity . *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW186 (April 2025), 33 pages. https://doi.org/10.1145/3711084

Authors' Contact Information: Tawanna R. Dillahunt, tdillahu@umich.edu, University of Michigan, School of Information, Ann Arbor, Michigan, USA; Kerby Shedden, kshedden@umich.edu, University of Michigan, Department of Statistics, Ann Arbor, Michigan, USA; Mila Ekaterina Filipof, mfilipof@umich.edu, University of Michigan, School of Information, Ann Arbor, Michigan, USA; Soyoung Lee, soyolee@umich.edu, University of Michigan, School of Information, Ann Arbor, Michigan, USA; Mustafa Naseem, mnaseem@umich.edu, University of Michigan, School of Information, Ann Arbor, Michigan, USA; Kentaro Toyama, toyama@umich.edu, University of Michigan, School of Information, Ann Arbor, Michigan, USA; Julie Hui, juliehui@umich.edu, University of Michigan, School of Information, Ann Arbor, Michigan, USA.

ACM 2573-0142/2025/4-ARTCSCW186
https://doi.org/10.1145/3711084

1 Introduction

Scholarship in human-computer interaction (HCI) and computer-supported cooperative work and social computing (CSCW) has extensively addressed digital inequality across three dimensions [78]:

- **Level 1:** Who has access to the Internet?
- **Level 2:** Who can navigate the Internet effectively?
- **Level 3:** How do people leverage resources on the Internet for economic and social mobility?

Addressing digital inequality in such a multi-dimensional way acknowledges the complex and wide spectrum of differences and disparities based on demographic factors such as age, race, gender, education, physical abilities, geographic location, household type, and level of economic development [25]. While this research has significantly advanced our understanding of digital inequality, it emphasizes individual capabilities. A focus on individual capacities overlooks how community-level factors such as social networks and collective digital competencies can influence digital engagement [43, 90, 94]. This oversight is especially concerning given that the coronavirus pandemic (COVID-19) further isolated individuals and exacerbated challenges around limited Internet access [9]. Our work challenges the existing framework by proposing a new construct—*community digital capacity*, which is distinct from averaging individuals' digital literacy. Instead, this construct encompasses community-level resources and networks. We define *community digital capacity* as the *extent to which the culture, infrastructure, and digital competence of family and community enable and support digital practices.*

Our work draws from HCI and CSCW researchers who have shown the need to understand the human or social infrastructure of information and communication technologies, particularly in developing regions [37, 69, 73, 77] and underserved areas. Such studies have highlighted the importance of understanding the existing 'village-like' infrastructure that already exists within underserved communities (e.g., [29]) and the notions of care (e.g., [57]) in these communities. Research has also shown that social networks can influence our motivation to learn, adopt, or resist new technologies, with people in our networks playing a crucial role in providing digital support and influencing our digital capacity [23, 27, 51], and that social capital and community collectives are essential for digital engagement in lower-income communities [27, 51].

Our work addresses the following questions: *How might we begin to conceptualize a measure to account for a community-level assessment of digital capacity in a reliable manner? What are the different dimensions of this measure?* This new approach to assessing digital readiness and resources at the community level includes the individual skills of community members, their supportive networks that facilitate digital engagement, and infrastructural resources. Our proposed metric aims to provide a more comprehensive understanding of a community's digital readiness and capacity to embrace digital opportunities. Our proposed metric is not designed to be a universally applicable measurement but one that accurately reflects the digital capacities of specific communities, particularly those that are underserved. This targeted focus helps identify and address specific barriers to digital access and utilization not typically covered by more generalized studies. While generalizability has its merits, particularly in studies involving larger, more accessible populations like college students or residents of countries with very high household internet penetration and educational attainment [86], our research intentionally shifts focus to address calls to include those often left at the margins by conventional digital literacy frameworks [92].

We operationalized an initial three-factor structure that consisted of the domains: individual digital capacity, social digital capacity, and infrastructure. Subsequent factor analyses reveal that our measure has two distinct dimensions: individual and social. To our knowledge, our proposed metric would be the first to measure community digital capacity, an essential instrument for under-resourced contexts. Our scale has significant practical implications. It provides policymakers, educators, and community leaders with a tool to enhance the effectiveness of digital inclusion efforts, ensuring that their interventions are tailored to meet their communities' unique needs. Specifically, our metric could serve as a foundational tool for designing inclusive digital policies and educational programs in domains such as healthcare, employment, and even the use of "smart city" technologies such as electric vehicles and "smart meters"—efforts aimed at improving the quality of life within communities. Therefore, our work extends digital inequality research and contributes to broader efforts aimed at enhancing our current assessment tools (e.g., [43, 72, 78, 93]), directly reaching out to communities traditionally underserved by society and strengthening community support systems that could enhance individual digital skills. With these interventions, such communities are better positioned to thrive in an increasingly digital world.

Our work makes theoretical and methodological contributions and is an important first step in understanding longitudinal changes over time within communities impacted by digital inequality. In addition to proposing a new construct, we provide theoretical reasoning for the importance of social digital capacity when considering efforts to bridge digital inequality. Social and communal aspects of digital inequality are conceptually meaningful and are about how supportive a community as a whole can be for individuals who could benefit from digital support.

2 Background and Related Work

In this section, we expand on the multi-dimensional nature of digital inequality and existing measures of digital inequality, articulating and justifying the need to focus on community-level resources and why HCI and CSCW are uniquely positioned to take on this call.

2.1 The Multi-dimensionality of Digital Inequality

The digital divide broadly encapsulates inequalities when it comes to the use and access of Information and Communication Technologies (ICTs), especially in political discourse [3]. While sometimes oversimplified as a dichotomy between those with internet access and those without [94], the digital divide is complex. A more nuanced understanding of the digital divide recognizes it as a multidimensional phenomenon with three levels of dimensions: first, in who has access to the Internet; second, who has skills to use the Internet; and third, how the Internet is used [24, 32, 43, 66, 78, 93].

2.1.1 Level 1 Digital Divide. The first level of the digital divide is about who has access to the Internet. It can include poor or no access to the internet or gaps in internet penetration [66]. It also refers to physical access to digital technologies. This level of the digital divide has traditionally been associated with sociodemographic factors such as geographic location, level of economic development, age, ethnicity, income, and education [66]. According to the comprehensive overview of the digital divide by Lythreatis et al., the first level has been a key focus in the literature when attempting to understand the basic disparities associated with the concept of the digital divide [66].

2.1.2 Level 2 Digital Divide. Digital literacy and digital and internet skills are terms often used to describe the second level of the digital divide. This level focuses on differences in patterns of Internet device use, knowledge, and digital skills [66]. Digital literacy refers to the basic skills needed for information processing and retrieval and digital media [56]. Broadly, digital literacy is related to the capacity and ability to use digital resources to manage access, integrate, analyze, and synthesize digital information [55]. Digital literacy encompasses a broad range of competencies related to digital technology use. Internet skills (a subset of digital literacy) refers to the competencies and abilities necessary to use the Internet effectively [89]. This distinction is important because internet use entails very specific tasks that require a specific skill set beyond general digital literacy such as online communication, information searching, and content creation [89]. The level two digital divide also includes psychological factors such as computer anxiety, low self-efficacy, and inequality of motivational access, all of which may prevent people from using specific technologies [92] are also associated with [66, 92].

2.1.3 Level 3 Digital Divide. The first two levels of the digital divide deal with utilizing and accessing ICTs. The third level focuses on disparities in how individuals leverage their digital skills to accomplish certain goals [66]. Stern et al., van Deursen et al., Van Deursen and Helsper note that despite having sufficient skills and technology, there are discrepancies in the benefits and consequences of their use of digital technologies and the Internet [84, 87, 88]. The third level of the digital divide encompasses consequences for individuals and society. In other words, it spans beyond disparities in one's ability to benefit personally and professionally from using technologies and how digital exclusion impacts broader societies at the social, economic, and political dimensions. The third level of the digital divide is the only level that extends beyond the individual. Our concept of the community digital capacity spans beyond the individual as well. Despite this similarity, we focus on the community level and derive our definition of community based on Cobigo et al. [17] review of prior definitions of the term in disciplines studying human interactions and behaviors. They defined community as "a group of people that interact and support each other, and are bounded by shared experiences or characteristics, a sense of belonging, and often by their physical proximity," [17, p.192]. Their analysis uncovered common themes such as physical proximity, shared values, beliefs, and worldviews [11], and belonging. Narrowing this definition and shared characteristics for our study, we define community as "a group of people with shared geographical location, culture, identity, and circumstance" [17].

The comprehensive review of the digital divide by Lythreatis et al. identifies new levels of the digital divide that have begun to emerge in recent literature: data inequalities (i.e., the unequal distribution of and the benefits derived from the use of data) and algorithmic awareness (i.e., the unequal access to knowledge about how algorithms buttress digital technologies) [66]. These emergent levels reinforce the importance of the third-level digital divide in considering the consequences and outcomes of digital use and ways to mitigate disparities in the outcomes of future digital inclusion efforts.

2.2 A Theoretical Model of Community Digital Capacity within HCI and CSCW

As described in Section 4.1, we conceptualize community digital capacity as the "extent to which the culture, infrastructure, and digital competence of family and community enable and support digital practices." Our work draws upon and extends the work of prominent scholars in the areas of digital inequality and technology adoption like Scheerder et al., Ragnedda and Muschert, Van Dijk, and Hargittai, [13, 43, 65, 72, 76, 78, 92, 93].

Policymakers and researchers use a combination of measures from these scholars to assess digital inequality comprehensively. Such measures are necessary for identifying specific areas of need, tracking progress over time, and informing strategies to effectively mitigate the digital divide.

2.2.1 Theoretical Foundation and Core Model Components. Our theoretical model acknowledges the nuances of the digital divide as more than a binary of access and non-access: it is a spectrum influenced by various socio-economic and cultural factors.

Individual Digital Capacity: Individual digital capacity represents basic digital skills and access at the individual level [92]. However, whereas all levels of the digital divide reference individual digital capacity—except the third level's focus on society—our model incorporates HCI and CSCW research insights that consider the role of shared activities and social resources to offset any limitations in an individual digital capacity.

Community Resources and Infrastructure: This area span beyond individual digital proficiency (e.g., [13]) to encompass community-level resources, networks, and infrastructure. Community support networks can mitigate individual digital literacy and digital resource access. For instance, a community center with reliable internet access and technical support can enhance overall digital capacity. This enhancement allows for social support networks to mitigate shortcomings in both individual digital literacy and digital resource access.

Social Support Networks: In general, access to peer or social networks provides access to new information, resources [19, 71], and even digital infrastructure (resources and accessibility). From the perspective of digital literacy, access to social networks could provide access to computer-related hardware or software, guidance, advice, and skills transfer [23, 27, 51]. For example, community members might assist each other with technical issues or share digital knowledge, fostering a supportive environment for digital engagement, influencing the level of individual digital capacity.

Exposure to others with varying beliefs about technology and digital literacy could also contribute to a deeper, more complex understanding of technology that goes beyond general usage. In order to activate digital engagement in lower-income communities, HCI researchers identified non-technical requirements that are needed such as social capital, social networks, and incubation from organizations [27]. Follow-on work also found that community collectives comprising resource-connecting organizations foster informal opportunities for digital engagement and troubleshooting [51].

2.2.2 Model Contributions. Despite the recognized influence of social networks on digital literacy [66] and the well-documented evidence of the qualitative benefits of community-based interventions to provide digital support [4, 6, 49, 50, 60, 62, 63], there are currently no quantitative measures to capture their effectiveness. Our work, which builds on our prior late-breaking work [30], addresses this gap. Our model considers practical questions such as: How many people could one call upon to ask a basic technology question? Do community members have help or someone to call on in case of a computer problem? Do they help others (e.g., friends and family) troubleshoot technology issues? Could an elderly community member call up a neighbor to ask for help setting up a Zoom call for a virtual appointment with a doctor or have a place where they could access reliable internet? Despite the recognized influence of social networks on digital literacy [66], explicit measures to assess community-level digital capacity remain scarce.

Our construct provides a framework that captures the complex interplay between individual skills, social support, and infrastructural resources. Building on prior research, this community perspective on digital literacy could provide a more nuanced understanding of the digital capabilities and gaps of under-resourced populations more likely to rely on social support to bridge gaps in basic resources. Along these lines, Kozubaev et al. similarly emphasize the need to understand existing relationships and practices within communities to ensure the successful and conscious implementation of smart technologies in communities like public housing [61]. In addition to contributing our theoretical model, we explain our scale development process and present the outcomes of two empirical studies to demonstrate the viability of our model.

3 Community Selection, IRB and Scale Development Process

3.1 Two Communities: Community 1 (PHO) and Community 2 (ARO)

We contrasted two distinct populations from southeastern Michigan—Communities 1 and 2—to assess the scale's meaningfulness and robustness across diverse groups. We chose these sites based on several criteria that align with the theoretical underpinnings of the measure. Community 1 is a public housing organization we will refer to as PHO. United States (U.S.) public housing communities encapsulate concentrated socioeconomic challenges and digital divides critical for testing our measure's comprehensiveness. PHO serves predominantly low-income African American families in Detroit. Our partnership enabled us to explore our measure within a densely populated urban area where communal resources and infrastructure might be pivotal in shaping digital engagement.

Over the past decade, Michigan has been within the top four U.S. states receiving the most refugees behind Texas, California, and New York [53]. Community 2 consisted of an Afghan refugee population and is referred to as ARO. It was also public housing; however, it is geographically dispersed within a 5-mile radius (versus PHO participants who were physically located in the same public housing development). This site was both a practical and opportunistic choice for validating our measures: practical because we had a prior relationship with the organization and unique because of the culturally distinct perspectives of its residents when compared to those in the first site. As a result our two communities represented distinct demographic and socio-economic profiles, offering diverse digital literacy levels and access to technology. Selecting public housing as our focal setting allows us to uncover differences in individual digital capacity, social connections, infrastructure, and possibly cultural norms that influence digital access and usage. This comparison is both opportunistic and strategic as it opens an opportunity to further strengthen and substantiate our central contribution: conceptualizing the measure and understanding whether such a measure is meaningful in underrepresented environments.

3.2 IRB Process

Our university's Institutional Review Board (IRB) reviewed the study protocol. The IRB granted an exemption from ongoing review, recognizing the low-risk nature of the research and our adherence to strict data protection standards. For each cognitive interview and for those opting to take online surveys in person, researchers explained the purpose of the research and obtained oral consent from these participants. Participants completing surveys online provided consent through click-through agreement.

We addressed literacy concerns particularly for some participants from Community 2 (ARO), who had very limited or no literacy skills. In these cases, our data enumerator read questions aloud from a tablet and entered responses on behalf of the participants to ensure accuracy and understanding. Again, oral consent was obtained, which adheres to best practices for overcoming literacy barriers as well documented in literature [40].

The research team adhered to best practices for data management to ensure the confidentiality and privacy of participants by anonymizing participant data and limiting data access to IRB-trained team members. We were also open about the goals, methods, and anticipated outcomes for the project and regularly communicated with stakeholders and participants to build trust and accountability.

3.3 Scale Development Process

Our scale development process followed a systematic nine-step procedure categorized into three overarching activities: 1) Item development, 2) Scale Development, and 3) Scale Validation. We followed established best practices in scale development literature [12, 67] similar to Hasan et al. in HCI [44]. Each step is briefly described below and further elaborated in subsequent sections:

Step 1: Item Development—The goal of this step is to refine and reduce survey items and ensure survey understandability among Community 1 participants.

(1) Construct Definition and Item Generation: Building on our related work, we defined the target construct, "Community Digital Capacity," and created an initial item pool of approximately 70 items to describe it.

(2) The research team and community partners from Community 1 reviewed each item to evaluate content validity to ensure item relevance and representativeness.
(3) We conducted cognitive interviews to pre-test questions for construct alignment and measurement validity.

Step 2: Development of the Scale and assessment with Community 1—The goal of this step is to explore which items loaded highly on each factor.

(1) We administered surveys to a representative sample of Community 1, the target population.
(2) Based on exploratory factor analysis, also described as late-breaking work [30], we removed items that did not correlate well with other items or failed to load significantly, reducing the item pool.
(3) Exploratory factor analysis helped to visualize and identify three latent factors within the scales, which we identified as "individual digital capacity," "social digital capacity," and "infrastructure."

Step 3: Scale Validation: Communities 1 and 2—This step aims to validate the scale.

(1) We performed confirmatory factor analysis with data from Communities 1 and 2 to see if the factor structure identified in Community 1 held in Community 2.
(2) We assessed the scale reliability through internal consistency analysis.
(3) We evaluated the criterion and construct validity through regression analysis.

We followed this rigorous procedure to ensure the development and validation of a robust measurement instrument.

4 Item Development

4.1 Construct Definition

Hasan et al. [74] clarified the need to define the construct before collecting the indicated items. Thus, following this suggestion, we defined our construct. We explain how we developed the initial set of items to characterize the construct. We aimed to propose a new construct—*community digital capacity*—a measure of digital literacy that spans beyond the individual.

There is no consensus on the definition of community. For the purposes of the study, we define community as "a group of people with shared geographical location, culture, identity, and circumstance," as suggested by Cobigo et al. [17]. Similarly, there is little consensus on a definition of digital literacy, and definitions vary based on various perspectives and agendas [81]. Comprehensively, Spires and Bartlett conceptualized digital literacy in three categories, which closely align with the levels of the digital divide described in the Related Work: 1) locating and utilizing digital content, 2) creating and curating digital content, and 3) dissemination of digital content [82]. We viewed digital capacity as a broader construct and were inspired by the term organizational digital capacity. Organizational digital capacity is how an organization's culture, policies, and infrastructure enable and encourage digital practices [58]. Costa et al. built on organizational digital capacity to conceptualize a school's digital capacity. Specifically, they described that capacity as how "culture, policies, infrastructure as well as digital competence of students and staff support the effective integration of technology in teaching and learning practices" [21, p.2].

Thus, our proposed measure is a multidimensional construct that encompasses the collective ability of a community to access, use, and benefit from digital technologies. It extends beyond individual digital literacy to include the shared resources, social support networks, and infrastructure that collectively influence digital engagement and proficiency within a community. As seen in prior HCI and CSCW literature, community digital capacity acknowledges that technology is embedded within the community's social fabric. Effective digital engagement depends on a supportive, resourceful, and connected environment [37, 69, 73, 77] and this notion continues to be supported by recent studies (e.g., [27, 29, 48, 57, 61]). However, there is currently no way for our field to begin measuring the effectiveness of our digital interventions or identify existing community strengths when assessing the types of support necessary in a community. Addressing these gaps in measurements is especially important given technology's role in supporting social determinants of health [7, 39], such as employment and economic mobility, transportation, education, housing and healthcare access, especially among marginalized communities.

Thus, building on Killen et al. and Costa et al. [21, 58], we conceptualize community digital capacity as the *"extent to which the culture, infrastructure, and digital competence of **family and community** enable and support digital practices."* When considering efforts to bridge digital inequality, communities, especially those experiencing marginalization, often play a pivotal initial role before educational or organizational interventions. We estimated an initial structure based on past research suggesting categories that community digital capacity might be divided:

(1) **Resources and accessibility:** The ability of members of the community to access digital devices and the Internet.
(2) **Individual digital capacity:** The magnitude and distribution of individual digital literacy within the community.
(3) **Social digital capacity:** The extent to which community members can receive and support digitally-mediated tasks. This capacity includes access to someone who could troubleshoot technology issues and the ability to access someone within their social network for technology-related assistance.
(4) **Infrastructure:** The physical and digital infrastructure available to the community, including access to the internet, digital devices, and community spaces that facilitate digital engagement.
(5) **Digital assets/currency:** The monetary assets in digital form that facilitate online transactions and financial operations (e.g., access to forms of electronic money and ways to leverage digital platforms to raise money outside of work).

By integrating these factors, our proposed measure of community digital capacity seeks to provide a more holistic understanding of digital literacy. This approach contributes to the theoretical discourse on digital inequality and offers practical insights for designing more effective digital inclusion initiatives. Our framework underscores the importance of considering the community as a critical unit of analysis when addressing the digital divide, moving beyond individual-centric measures to a more inclusive and comprehensive understanding of digital literacy.

4.2 Initial Item Pool Generation

The research team led the initial development of item collection from September 2021 to May 2022. We took a mixed methods approach to enhance the reliability, validity, and applicability of our proposed measure. Our approach, similar to Hasan et al. [44], began by deriving many items from prior literature, as described in Section 2.2. We also created additional items independently before leveraging similar questions from existing instruments. In some cases, we modified existing questions to fit our setting. To ensure the content validity and relevance of our measures, we followed a systematic refinement process that involved the following steps: selection and adaption of established scales, community partner feedback, pilot testing, and item modification.

We started with approximately 70 items. While we adapted most items from resources and accessibility and individual digital capacity from existing scales, we created our own questions about social digital capacity, digital assets, and currency. Borrowing from Read et al., we inquired about smartphone instead of cellphone access. The original question was based on a single measurement to assess cellphone access, " 1) I own or regularly use a cellphone; 2) I have NOT owned or regularly used a cellphone in the past; 3) I do not own or regularly use a cellphone now, BUT I have IN THE PAST YEAR [75]." We discussed as a group whether to use the term "cellphone" or "smartphone." We decided to use "smartphone" based on prior research, which indicated a preference for "smartphone." We defined a smartphone as, "a mobile phone (such as iPhone, Samsung Android) that could be used for internet browsing and multiple functions such as music, game, camera, videos, social media on a touch screen in addition to voice call and text messaging." We relied upon Boot et al.'s "Computer Proficiency Questionnaire", or CPQ, to assess individual digital capacity. CPQ was developed to measure computer proficiency among people with a range of proficiencies, from non-computer users to extremely skilled users. The measure includes assessments ranging from computer basics, printing, and the Internet to calendering software and multimedia use [13]. We aimed to select at least two skill assessments ranging from easy (i.e., can send a text message or use a keyboard) to difficult (send an email to multiple people at the same time, use a spreadsheet). The research team determined these two skills with our community partner. We also modified the scales to "I can do this easily," "I can do this, but it can be difficult," "I don't know how to do this," and "I don't understand the statement / I have never tried," so that we could differentiate among responses and apply a score to help rank digital proficiency/capacity.

To determine questions for social digital capacity, we were inspired by Cohen and Hoberman's perceived availability of social support measure (ISEL) [18]. Items relating to "knowing someone willing to help" were adapted from these scales. We wanted to understand whether people could borrow devices from others if they did not own them or whether or not they knew someone who could teach them tasks they needed to learn. We modified the type and frequency of support provided. For instance, "I know someone who is willing to help me *once in a while* if I have difficulty doing things on my smartphone." We considered tangible support in addition to skills-based support. For instance, "If I lost access to my own devices and needed to use the Internet, I know someone who would loan me their devices." We also adapted questions from the Internet Social Capital Scales (ISCS)' Bonding Social Capital Scale [95], which extended the ISEL measure, including name, position, and resource generators as indexes to accumulate scores. For example, we included a question for respondents to list between 0 and 5 names, indicating whether they are a family member, close friend, acquaintance, or professional whose job is to help them with their computer devices. We drew inspiration from international indices like the ICT Development Index [5] infrastructure and access indicators for our infrastructure-related questions. Our goal was to include questions about the availability and reliability of the internet and access outside the home. We also recognized that for any given community, there was a likelihood of seeking and validating data from publicly available sources or applying the national index for specific communities. Finally, we concluded with questions related to digital assets. Such questions were inspired by prior HCI and CSCW research of technology within the context of marginalized communities and included questions regarding the use of social media, number of friends or followers, whether people sought online credentialing or obtained an online degree, whether they would feel comfortable posting a request for help or resources online, and the greatest amount of funds raised at one time using digital apps or websites. The survey concluded with demographic information.

Similar to Hasan et al., we followed a bottom-up approach. We began with an inclusive set of items that were not narrowly scoped, which often concludes with a multifactor situation [44]. Similarly, rather than aiming to encompass predetermined contexts or traits, all subsequent decisions to remove items were based exclusively on their fit with the theoretical framework and statistical criteria [44]. We began with approximately 70 items, exceeding the twofold expected amount of our final scale of 20 items, which satisfies Kline, Schinka and Velicer's recommended criteria.

4.3 Refinement of Items

The research team met regularly to discuss items before presenting the initial item set to our community partners for feedback. Points of contention included whether to use subjective framing such as "confidence," which, while appropriate for our setting, might not span beyond HCI or CSCW communities, and whether to keep the fourth proposed factor of digital assets/currency. Therefore, following Hasan et al., we were careful to ensure items necessitated only a minimal level of respondents' subjective judgment and that items were impartial [20].

We also discussed the desire for the instrument to capture a rough sense of community digital capacity, assuming responses from a representative sample of any community of interest. We anticipated that answers to the questions would measurably shift over time as a result of future interventions aimed at bridging the digital divide. Ideally, this measure would be used to assess such shifts. We imagined a way to convert responses to the questionnaire to a single number so that it could become an index in the future. Thus, a community could *hypothetically* have a community digital capacity score of 6.4 to 7.2 within two years of introducing an intervention to bridge the digital divide.

In some cases, some questions did not have an obvious conversion score, and we aimed to clarify such questions. At the end of these early discussions, we had about 43 items in our pool.

4.4 Evaluating Content Validity

We evaluated content validity to ensure item relevance and representativeness with the support of community partners. The university researchers reviewed the initial survey with community partners from public housing to clarify, add, and eliminate unnecessary questions. We aimed to keep the survey completion time to approximately 10-15 minutes.

Using the framework of common errors identified in cognitive interviews [2] (comprehension, retrieval, estimation/judgment, and reporting) we modified our survey accordingly. In some cases, we needed to clarify our survey instructions (e.g., moving direction text such as "check all that apply" to the beginning of questions (Q1, Q3) and integrating the definition of a smartphone into the question text). For instance, we added "flip/feature" phones to describe what smartphones were not. We also clarified and changed words interpreted in multiple ways. For instance, we changed the most "powerful" person to the most influential person (Q12). We also separated question items in our matrix table into sub-matrix questions to increase legibility and recognition (Q4 and Q6). To address retrieval issues, we added detailed examples to remind respondents of potential responses and clarify the intention of questions. For instance, we added categorical examples of who the most influential person might be (e.g., a school official, community, community or political leader, celebrity, etc.). We adjusted attitude response scales that were hard to distinguish to alleviate estimation/judgment-based errors (i.e., combined, "I don't know how to do this" to "I have never tried" (Q5). Finally, we added open response forms to address reporting errors to formulate more detailed answers and reasoning. Again, to understand the most influential person, we added an open response for respondents to include the person's title/position and affiliation and explain why the person was influential. At the end of our discussions with our community partner, our final list included 51 items before pre-testing with community members.

4.5 Pre-test Questions

Next, we conducted an iterative series of cognitive interviews to identify and correct problems with our survey questions [8, 85].[1] The cognitive interviews aimed to ensure our questions were understandable (content validity) and determine whether our questions helped us measure what

[1]Study protocols are included as auxiliary material.

we intended to measure (face validity). Questions from our cognitive interviews ranged from interviewee's feelings about the overall length of the survey and how they decided to answer questions the way they did, to requests about what they thought questions were asking in their own words. We also sought clarification about frequency—"Can you explain your understanding (about how many times) of 'Multiple times a week' but not 'Everyday'?"

We identified groups of people most relevant to the study until we yielded relatively few new insights [8, 85]. In doing so we followed the best practices of Beatty and Willis, Terry and Fobia regarding the dominant paradigms and key decisions about cognitive interview study design. These authors state that sample sizes in cognitive interviewing may often fall short of capturing all insights, yet the principle of diminishing returns is crucial. This approach indicates that researchers should continue interviews until new insights diminish, often requiring a flexible number of participants. Therefore, we piloted our survey with four family members of one of the researchers. We followed these think-aloud interviews with six more respondents from our population of focus. While no universal standard exists, our use of 10 participants, including initial feedback and cognitive interviews, was designed to balance thoroughness with practical constraints. Those completing our pilot surveys were not paid; however, we paid community members from our population of focus $20 for each interview. Interviews spanned from 27 minutes to 1 hour and 40 minutes. On average, each interview lasted 85 minutes or 1 hour and 25 minutes (1.15 hours). With guidance from a survey expert, we coded responses for "question comprehension, recall, and judgment."

Our pilot cognitive interview results helped us identify questions that were difficult to understand, questions that required clarification (comprehension), and those that were candidates for removal. We received feedback that the question, "Who is the most powerful person you have connected with online," was unclear. There was a bit of uncertainty about what constituted something occurring "online." For example, "Does online mean Zoom?" "Do you mean connected through the Internet? What about messenger?" The initial think-alouds also revealed that respondents had difficulty recalling questions inquiring about the number of Facebook friends or social media followers they had on specific sites. Thus, we dropped these types of questions. For unclear questions, respondents suggested including parenthetical examples. Questions about resources and digital assets/currency eventually dropped, resulting in 29 items in total: 15 individual digital capacity items (3 Likert scale), 10 social digital capacity items (True/False), and 4 infrastructure items (True/False) (See Figure 1). Finally, as anticipated, pilot participants identified typos and suggestions for making our survey more consistent.

5 Development of the Scale

This section outlines our dissemination process to collect our data and our factor analyses. Because we collaborated with a non-profit housing and refugee support organization, we included a question to clarify where residents resided. We used both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), two main types of analyses based on the common factor model [54], as a part of our methodological approach for complementary purposes in the development and validation of our scale. EFA is an exploratory technique that determines the appropriate number of common factors and identifies which measured variables effectively represent underlying latent dimensions [15]. EFA is used in the early stages of scale creation and construct validation, while CFA is used in later phases of scale development after establishing the underlying structure [15]. CFA allowed us to validate the construct and measurement invariance and examine the latent structure of our instrument [15]. Designing our approach for iterative validation of our instrument, CFA, with the Afghan refugee organization's data allowed us to validate the scale across distinct groups [15], specifically in terms of culture, race, and context. After we administered the survey at the public housing community and conducted an EFA, we engaged in the item reduction process described in

Original name	# on Plot	Renamed variable	Question
		a_	**The next set of statements asks about your ability to do different tasks with a computer, smartphone, or flip phone/feature phone. Choose how easy or difficult it is for you to do each task:**
Q5-1_1	1	a_text	Send a text message
Q5-1_2	2	a_keyboard	Use a keyboard (including one on a touchscreen)
Q5-1_3	3	a_download	Download an app
Q5-1_4	4	a_watch	Watch videos (movies)
Q5-1_5	5	a_record	Record a video
Q5-1_6	6	a_openread	Open and read emails
Q5-1_7	7	a_email	Send the same email to multiple people at the same time
Q5-2_1	8	a_search	Search for information online (example: using Google)
Q5-2_2	9	a_calendar	Enter events and appointments into a digital calendar (example: on Google, Outlook or Apple Calendar)
Q5-2_3	10	a_videocall	Join a video call (example: on Zoom, Google Meet or FaceTime)
Q5-2_4	11	a_payment	Make a digital payment to a friend (example: using Cash App, Paypal, Venmo, Zelle or your bank)
Q5-2_5	12	a_purchase	Purchase an item online
Q5-3_1	13	a_socialmedia	Add, follow or accept new friends/connections on social media like Facebook, Instagram or Twitter
Q5-3_2	14	a_prvsetting	Change my privacy settings on social media like Facebook, Instagram or Twitter so that only my friends can see my posts
Q5-3_3	15	*a_spreadsheet	Use a spreadsheet like Excel/Google sheets to keep track of things such as a contact list, items for tax preparation or to do a budget
		*internet_	**The next statements ask about Internet accessibility. Choose True or False for each statement:**
Q8_1	16	*internet_address	Internet service is available at my address (example: AT&T, Xfinity by Comcast)
Q8_2	17	*internet_reliable	There is reliable Internet connection where I live (example: I can reliably hold a zoom meeting, I can watch a video without it freezing)
Q8_3	18	*internet_place	Over the past year, most of my time was spent in places that had Internet access.
Q8_4	19	*internet_outside	I am able to access the Internet on a smartphone or flip phone/feature phone outside of my home (example: at the store, on the bus)
		know_	**If I have difficulty doing something on my computer, smartphone, or flip phone/feature phone:**
Q6-1_1	20	know_once	I know someone who is willing to help me once in a while.
Q6-1_2	21	know_daily	I know someone who is willing to help me on a daily basis.
Q6-2_1	22	**know_prob	New variable combined from know_few prob and know_allprob (few_prob: I know someone who can solve - a few of the problems I have with my computer, smartphone, or flip phone/feature phone; know_allprob "I know someone who can solve almost all of the problems I have with my computer, smartphone, or flip phone/feature phone)
		know_	**I know someone:**
Q6-3_1	23	know_teach	Who would - patiently teach me how to learn a new skill on a computer, smartphone, or flip phone/feature phone.
Q6-3_2	24	know_anytask	Who would - do any task that requires a computer, smartphone, or flip phone/feature phone on my behalf.
Q6-3_3	25	know_provide	Who would - provide me with a computer if I needed it for my job or school.
		know_	**If I lost access to my computer, smartphone, or flip phone/feature phone and needed to use the Internet:**
Q6-4_1	26	know_loan	I know someone who would loan me their computer, smartphone, or flip phone/feature phone.
Q6-4_2	27	know_comeover	I know someone who would let me come over to their place and use their computer, smartphone, or flip phone/feature phone.
Q6-4_3	28	know_place	I know a place where I could go and affordably use the Internet.

Fig. 1. List of 29 Items before Item Reduction (15 individual digital capacity items, represented by variable *a_*, 10 social digital capacity items, represented by variable *know_*, and 4 infrastructure items, represented by *internet_*). "Original name" corresponds to the final survey questions[2] and the # on Plot align with EFA results in Figures 2 and 3. * - Items that were eventually dropped. ** - Collapsed from two items.

Subsections 5.1.3 and 6.1. After conducting an EFA and CFA with the public housing community, we repeated the process with the Afghan refugee community. This is the conventional way to develop hypotheses and assess whether the hypotheses hold across cohorts [52].

5.1 Administer Surveys

We administered surveys at a public housing organization, referred to as PHO, serving predominantly low-income African American families in Detroit. For context, approximately 25 to 30% of Detroiters have no internet access including no access to cellular data [68]. At the time of the study, there were increasing city-wide efforts to distribute computer devices [36, 68], focusing on Level 1 Digital Divide challenges. We compensated participants $10 for providing valid and complete survey responses.

5.1.1 Public Housing Organization (PHO). Collaborating with our public housing community partner, we employed multiple offline and online approaches and purposive sampling strategies to collect data. Our initial online survey was distributed between April 25 and May 6, 2022. However, we paused our online survey collection because results showed that current residents were not completing it online. Thus, we restarted data collection a month later (from June 8 - July 26), employing a snowball sampling approach. We held several in-person survey sessions with printed surveys and allowed respondents to complete them online. In this case, residents became seed participants, helping to spread the survey within their networks. While not random, this method was necessary to reach effectively a harder-to-reach population [96].

Offline approaches included in-person attempts to obtain responses at offline community events and door-to-door. Our community partner further supported this effort by sending survey links via SMS, targeting community members who consented to receive their text messages. The average time to complete the survey for respondents completing more than 80% of the survey was 3.59 hours. However, some respondents left the survey and returned at a later time to complete it. Thus, this time is not an accurate account of the average time to completion. Most respondents took less than 30 minutes to complete the survey.

The survey was initially targeted at specific community members, but it was inadvertently shared in online Facebook groups, which led to it reaching a broader audience. As a result, we received a total of 721 responses. We excluded responses that suggested low engagement such as "speeders" (any survey time to completion less than 100 seconds), "straight lines" (respondents who entered the same answer multiple times), "slackers" (those who completed less than 80% of the survey), and survey bots [10]. We also excluded responses indicating potentially false data, such as invalid zip codes or zip codes that did not align with respondents' earlier claims of living in the public housing community. We detected survey bots primarily from invalid open-ended responses. These exclusions were critical in maintaining the integrity of our purposive sample.

We determined respondents' locations based on self-reported zip codes. Using Tableau, one of the authors plotted zip codes onto a map of the United States. From there, they selected responses from the state and filtered out city-based zip codes in an Excel spreadsheet. We filtered respondents from our community partners based on the specific zip code of the community development. We removed respondents with fewer than 20 of our 29 items completed. We kept duplicate IP addresses for respondents from the public housing site as we entered in-person surveys from the same laptop, and there was a computer lab for participants to complete surveys. In total, there were 489 valid responses. One hundred six (N=106) respondents were from Michigan, 383 were non-Michiganders, and only 56 valid responses were from the partnering community. A total of 87 Detroiters completed the survey. We focus on the 56 valid responses from our public housing partner because our study focuses on a specific geographic community.

5.1.2 Participant Demographics. Out of the 56 PHO survey respondents completing demographic information, the average age was 46 with an age range of 19 to 76 years (SD=13). The majority of respondents were women. Of those responding, 63% were women and 29% men. One person was non-binary, and another chose to self-describe. Of 61 respondents, the majority (32%) earned less than $10k (combined income of all household members in the previous year), 29% earned between $40,001 - $80k, and 21% earned between $10,0001 - $20k per year. The majority of respondents were Black (71%) and the remaining were White (29%) (See Table 4 for a summary). Most respondents (32%) were high school graduates with a diploma or the equivalent, (25%) completed some college with no degree, and (20%) completed an associate or bachelor's degree and reported being full-time employees (37%) (working 40+ hours a week); (13%) noted being a homemaker. Of the six individuals who reported being unable to work due to disabilities, ambulatory, cognitive, self-care, and independent living difficulties were noted.

5.1.3 Item Reduction. The initial results of our proposed measure of community digital capacity were promising. While prior work focused on assessing the skills component of digital literacy, the results of our survey revealed the various dimensions that constitute digital literacy. We use the goodness of fit, internal consistency, and uni/multidimensionality to describe aspects of the proposed community digital capacity measure. A single item, "Most of my time was spent in places that had Internet access," did not load with the other infrastructure items. Thus, we excluded this item from the proceeding confirmatory factor analysis, which was informed by the results of the exploratory analysis. We combined know_fewprob and know_allprob_ (i.e., to create Q6-2_1), given inconsistent responses. If a respondent knew anyone who could address any problem, we combined it as '1' versus differentiating between knowing someone who could solve any problem versus someone who could solve all problems. We created a new three-level variable, know_prob, that could account for neither, one, or both (i.e., values 1, 2, or 3). Thus, our final survey item count was 28 instead of 29.

5.1.4 Factor Structure Identification. We then used exploratory factor analysis to help visualize the differences between the two factors. We graphically displayed factor analysis results as a plot whose axes correspond to loadings of the first (individual digital capacity) and second (social digital capacity) estimated factors for each item (See Figure 2). Three colors were used to distinguish the three domains of items: orange lines each represent an item from our social digital capacity dimension, green lines Individual digital capacity, and blue lines Infrastructure. It is clear from the figure that the social and individual digital capacity items form independent common factors that are orthogonal to each other. While these results are promising, this was not the case for infrastructure items (several of these items are difficult to see as they overlap with individual digital capacity items).

6 Scale Validation

We confirmed our hypothesized factor structure by conducting a confirmatory factor analysis across PHO and an EFA and CFA across our second site, ARO. We include regression analysis to give us insight into what these constructs mean and to confirm alignment with prior research (i.e., nomological validity [42]).

6.1 Confirmatory Factor Analysis of Community 1: PHO

By convention, using the standardized root mean square residual (SRMR) statistic, SRMR < 0.1 is considered a good fit. Calculating SRMR for two factors with all items in the EFA model across PHO resulted in SRMR=.096, which was acceptable; however, there was no structure to the loadings (See Appendix C), showing that none of the loadings were zero. So, while we achieved a good fit

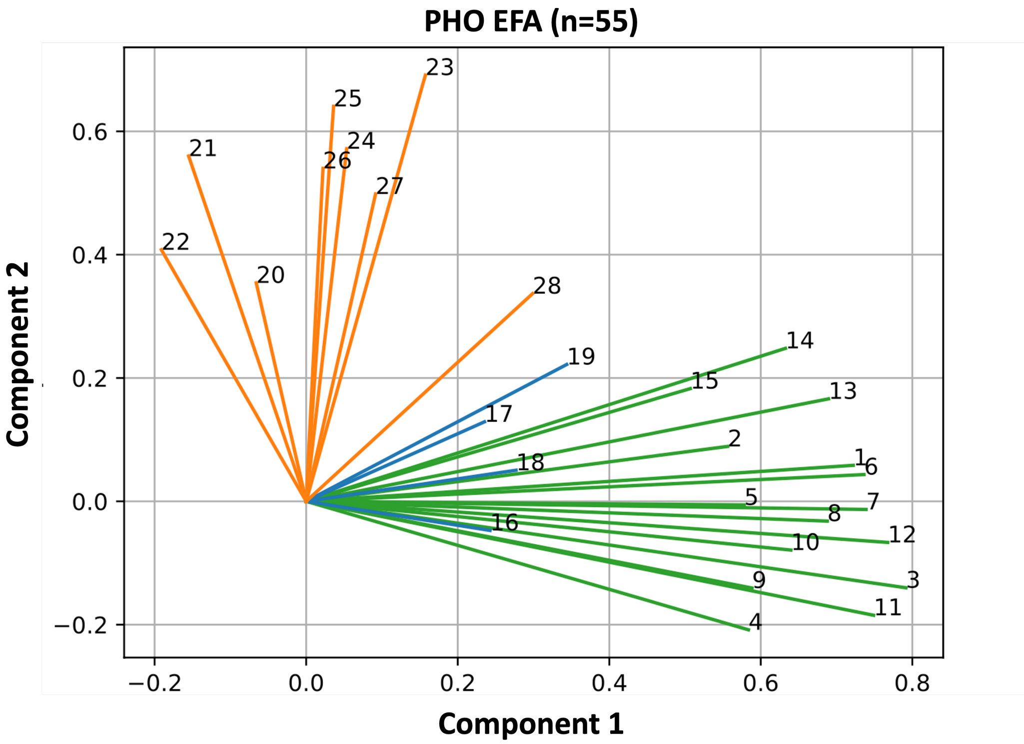

Fig. 2. Graphical Display of PHO EFA (Each color represents a single factor - Orange: Social digital capacity; Blue: Infrastructure and Green: Individual digital capacity).

with all loadings, our results were harder to interpret. Thus, we switched to our CFA model (see Table 1) to assess goodness of fit, removing an additional 8 items (a_spreadsheet, k_place, k_prob, k_daily, i_reliable, i_address, i_outside, and i_place) and bringing the final item count down to 20. This resulted in an SRMR=.099. The positive coefficients led to a more interpretable model structure, where one factor represents individual digital capacity, or (a_)bility items, and the second factor represents social digital capacity, or know_ items. The loadings on these 20 items nearly correspond to equally weighted averages of the ability and knowledge items for the two factors, respectively. The resulting two factors are almost uncorrelated ($r \sim 0.1$). Thus, we could conclude that in the PHO data, the ability and knowledge items (except those dropped as described above) can be summarized as simple averages. This means that the items within each category function similarly, allowing us to sum totals or calculate an average without needing to drop items or assign complex weights. This approach simplifies the analysis and indicates that our items can be treated as a cohesive and interchangeable set, enhancing the measure's practicality and interpretability.

Results from Table 1 (PHO) suggest that most Factor 1 (individual digital capacity or ability items) from the first community, PHO, have significant positive loadings, ranging from 0.55 (a_keyboard) to 0.80 (a_download). Factor 2 items (social digital capacity, or (k)now_ items) have much lower loadings (the highest being k_provide at .68).

Item	f1	f2	srmr
a_text	0.74	0.00	0.119
a_keyboard	0.55	0.00	0.098
a_download	0.80	0.00	0.063
a_watch	0.61	0.00	0.110
a_record	0.57	0.00	0.098
a_openread	0.75	0.00	0.097
a_email	0.72	0.00	0.078
a_search	0.70	0.00	0.091
a_calendar	0.57	0.00	0.115
a_videocall	0.63	0.00	0.094
a_payment	0.74	0.00	0.084
a_purchase	0.77	0.00	0.099
a_socialmedia	0.69	0.00	0.110
a_prvsetting	0.62	0.00	0.130
k_once	0.00	0.29	0.083
k_teach	0.00	0.65	0.129
k_anytask	0.00	0.50	0.113
k_provide	0.00	0.68	0.094
k_loan	0.00	0.62	0.092
k_comeover	0.00	0.57	0.109

Table 1. PHO Confirmatory Factory (CFA) Loadings: SRMR=0.099. Loadings were rounded to the nearest two decimal places and SRMR to the nearest three.

6.2 Pilot Survey

We systematically administered surveys across a second site to ensure a comprehensive understanding and validation of our proposed measure. The second site, which we will refer to as the Afghan Refugee Organization (ARO), is a resettlement agency in Metro-Detroit in a neighboring county that has helped resettle thousands of people from around the world to the United States. The agency offers programs to assist refugees and their families with social services necessary for a smooth transition to their new community and help them strive for self-sufficiency. Many of ARO staff are bilingual and bi-cultural professionals. While participants from PHO were physically located in the same public housing development, ARO participants were geographically dispersed within a 5-mile radius. We compensated participants $10 for providing valid and complete survey responses.

6.2.1 Afghan Refugee Organization (ARO). Prior CSCW work highlights the importance of sociocultural and sociotechnical adaptation and their relationship to migrants' distinct use of technologies, tools, skills, and willingness to use technologies. This is especially relevant for our population community's digital capacity. Afghanistan is home to several major ethnic groups with distinct languages, traditions, and gender norms that vary by region. Our goal, however, was not to identify the level of digital literacy among Afghan refugees but to validate our measure across this population and determine its meaningfulness. By understanding these communities' unique characteristics and challenges, we can assess the applicability and relevance of our digital capacity framework in diverse contexts.

Collaborating with our Afghan refugee organizational partner, we translated our surveys into Dari and Pashto. The translation process involved a review by internal staff and Afghan refugees who were actively engaged with the organization to ensure cultural and linguistic appropriateness. Based on their feedback, we removed two questions about posting a request for help or resources and raising funds online to ensure cultural sensitivity, which was crucial. All questions constituting our proposed measure of community digital capacity remained intact.

The organization advertised the survey offline at their site. Availability to complete the survey was also advertised via WhatsApp[3] groups. A community organization staff member provided one-on-one support to complete the survey on tablets at their homes or ARO. Due to literacy challenges, many women verbally provided their responses to the surveys, which their husbands then entered into the tablets. This approach was necessary to ensure all voices within the community were represented in the survey. However, it introduces a layer of second-hand reporting that must be acknowledged in the analysis.

The survey was administered from April 24, 2023 - June 12, 2023. The first and second sets of surveys were administered in Pashto and Dari, respectively, ensuring linguistic inclusivity. Most respondents took less than 10 minutes to complete the survey[4]. The initial survey yielded 108 responses. We excluded responses that suggested low engagement or potential false data, following the same procedure as with PHO. However, since everyone who completed the survey was from our anticipated population, no "slackers" or survey bots were detected. We removed respondents with fewer than 20 of our 29 items completed. We kept duplicate IP addresses for respondents from the site (given that surveys were completed in the same household or tablets). A total of 108 Afghan refugees completed the survey, with 97 valid responses. Of these valid responses, 77 completed the survey in Pashto, and 20 completed the survey in Dari. This ratio reflected the demographic ratio of Pashto to Dari-speaking residents in the region.

6.2.2 Participant Demographics. Out of the 97 respondents who completed demographic information, the average age was 31, 15 years younger than our PHO demographic, with an age range of 19 to 56 years (SD=8.5). The majority (59%) of respondents were men. Of the 96 respondents, (55%) earned less than $10k (combined income of all household members in the previous year), and 34% preferred not to say or did not know. The majority of respondents (36%) completed schooling from nursery school to 8th grade, (31%) did not complete any schooling, and (20%) completed some high school but did not receive a diploma. Most individuals (51%) reported being full-time employees (working 40+ hours a week) and (33%) noted being a homemaker. See Tables 4 and Table 5 in Appendix B for a summary of respondent demographic characteristics across both cohorts.

6.3 Exploratory and Confirmatory Factor Analyses Community 2: ARO

Aligning with our PHO EFA results, Figure 3 shows that the social and individual digital capacity items (orange and green lines) form independent common factors that are orthogonal to each other (the exception is the social item 25 or know_provide). Respondents had concordant responses in 20-24 (know_once, know_daily, know_prob, know_teach, know_anytask) and 26-28 (know_loan, know_comeover, know_place) but there was discordance between the two question sets. In other words, people who responded false to items 20-24 tended to respond true to items 26-28. We also see the Infrastructure factor failing to load. The two samples are somewhat concordant, though imperfect, when comparing Figures 2 and 3. While all items loaded concordantly across Factor 1 for

[3] According to the National Resource Center for Refugee, Immigrants, and Migrants (NRC-RIM), the WhatsApp platform is the preferred platform for refugees, immigrants, and migrants, offering an easy-to-use communication platform with people domestically and abroad [1].

[4] We believe ARO respondents finished the survey in 10 minutes versus 30 for PHO given the in-person support versus some PHO residents completing the survey only. We also believe that because female participants answered after their partner, they might have responded faster because any misunderstandings had been previously clarified.

ARO the loading was not as tight as PHO. Factor 2 items for ARO were more scattered; however, they all point in the same direction and form a common factor. The scattered items might suggest the different social context of ARO.

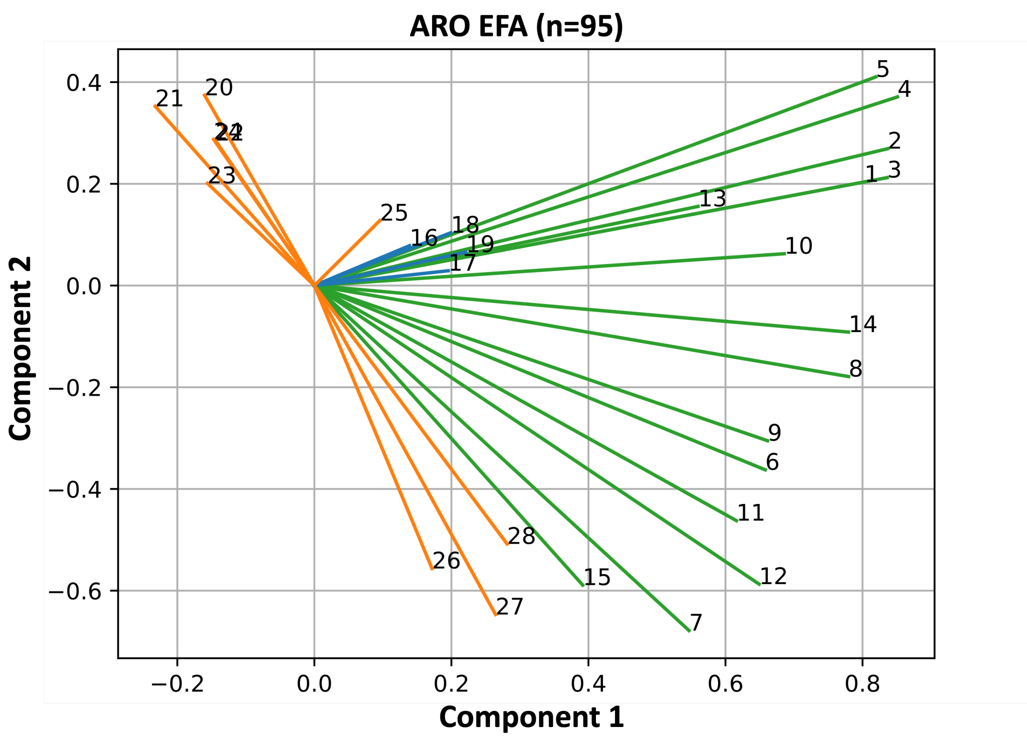

Fig. 3. Graphical Display of ARO EFA (Each color represents a single factor - Orange: Social digital capacity and Green: Individual digital capacity).

Next, we conducted a CFA (Table 2), considering whether this same factor structure fit the ARO data (this would be the standard approach when developing scales on the PHO cohort and evaluating it on the ARO cohort). When comparing PHO results to the ARO community, we see similarities for Factor 1. The loadings for Factor 1 (i.e., individual digital capacity) are higher in ARO, indicating a strong association between individual digital capacity and the corresponding ability items. However, we see differences in Factor 2 loadings (social digital capacity), suggesting a more complex relationship, again possibly suggesting contextual and cultural differences. Unfortunately, the SRMR of this model for the ARO data was too high: SRMR=0.16 (See Table 2 and Appendix C for details). Thus, quantitatively, our model does not align with the PHO data.

Based on these results, we see negative loadings on both k_loan and k_comeover, which suggests that people with higher social digital capacity may be less likely to know someone who could loan them a digital device (in the case of k_loan) or less likely to know someone who would let them come over to their place and use their digital device. These results could also indicate that the higher someone's social digital capacity, the less reliant they are on this type of in-person assistance. We later discuss possible qualitative alignments with our findings, given the contextual differences between the two sites and their implications. To aid our discussion, we include a table of cohort means across each item in Table 3 of Appendix A. Overall PHO means are generally higher in comparison to ARO. However, a_openread, a_email, a_purchase, and a_spreadsheet are much higher in PHO compared to ARO, and ARO has lower means for social items like k_place, k_comeover, and k_loan.

Item	f1	f2
a_text	0.828488	0.000000
a_keyboard	0.870202	0.000000
a_download	0.848218	0.000000
a_watch	0.859800	0.000000
a_record	0.832164	0.000000
a_openread	0.610814	0.000000
a_email	0.443216	0.000000
a_search	0.779571	0.000000
a_calendar	0.636950	0.000000
a_videocall	0.694264	0.000000
a_payment	0.549293	0.000000
a_purchase	0.559409	0.000000
a_socialmedia	0.585724	0.000000
a_prvsetting	0.785122	0.000000
k_once	0.000000	0.194026
k_teach	0.000000	0.801743
k_anytask	0.000000	0.801298
k_provide	0.000000	0.619580
k_loan	0.000000	-0.205844
k_comeover	0.000000	-0.234596

Table 2. ARO Confirmatory Factor Analysis Loadings, SRMR=0.162

6.4 Reliability Tests

While Cronbach's alpha is a common approach to internal consistency and assessing scale reliability and validity, it has limitations in its underlying assumptions of equal item loadings and can be misleading as it does not account well for multidimensionality, oversimplifying scale structure [16, 41, 80, 83]. For scales that capture multiple underlying constructs, Cronbach's alpha can lead to misleading reliability estimates [16, 41, 80, 83], while CFA handles multidimensionality [14]. Using SRMR as a fit index in CFA accurately represents the scale's reliability [47]. Quantitatively, our CFA/SRMR results for PHO show that all items on one factor form a coherent scale, demonstrating that CFA/SRMR is a more suitable and robust approach for this work.

6.5 Regressions on EFA and CFA

We conducted two sets of ordinary least squares (OLS) regression analyses (See Appendix D): (1) basic models including only age and sex and (2) more extensive regression models for PHO that included additional variables like household size. Sample sizes in different regression models differed based on missing covariate data. The primary purpose of presenting regression analyses is to uncover how factors relate to variables of interest like age, gender, employment, or household size and evaluate the measure's validity by assessing how well the results aligned with prior research.

We conducted regressions following EFA to explore patterns in our data and regressions following CFA to test whether our results aligned with established hypotheses related to measures of digital literacy (e.g., [91]). Overall, key findings regarding covariate effects included that (1) education predicts for higher individual digital capacity and (2) work predicts for lower individual digital capacity in PHO. We find that for ARO, (3) Pashto speaking predicts for lower individual digital capacity, being a (4) man and (5) Pashto speaking predicts for lower social digital capacity, and (5) age predicts for higher social digital capacity. We have provided the results of our regressions in Tables 11 - 14 in the Appendix and discuss the implications of their results next.

7 Discussion

Based on prior CSCW and HCI for Development literature, we propose the construct *community digital capacity*, an instrument unique in assessing community digital capacity and distinct from aggregating individual digital literacies, offering a novel measure for researchers and practitioners to assess access to shared digital resources and support. Our proposed measure consists of 20 items across two distinct dimensions: *individual digital capacity* (14 items) and *social digital capacity* (6 items). These dimensions were revealed through exploratory and confirmatory factor analyses. *Community Digital Capacity* highlights the overall supportiveness of communities in the context of digital literacy and can potentially compensate for or enhance individual digital literacy, as suggested by prior CSCW work (e.g., [27, 29, 48, 63]). We empirically tested our measure's factor structure across two distinct Southeastern Michigan populations. All social items were orthogonal to individual items in PHO, confirming our rationale for this work and suggesting the survey's design is naturally followed.

As stated, we turned to regressions following CFA to further test our results alongside established hypotheses. Our PHO regression analyses uncovered a single covariate effect that aligns with past research [91]. Education is associated with higher individual capacity. Prior digital divide research frequently observes that positional categorical inequalities like income, work, gender, and age tend to predict higher individual capacity [91]. However, income did not predict for higher individual capacity across either cohort, likely given that respondents were from lower-economic communities. Along a similar line, in PHO, work predicts for lower individual capacity. We believe that this finding is because the most common occupations of those living in public housing tend to pay lower wages (cashiers, nursing, home health aides, personal home care aides, and janitors are among the most common occupations) [70]. Such occupations are uncommonly associated with extensive technology use. In the future, we will be sure to collect specific occupational data. Age in ARO, predicted for *higher* digital capacity, signaling that younger people had lower digital literacy, countering prior digital literacy trends [91] and perhaps suggesting cultural nuances. As a reminder, ARO's average age was 31, 15 years younger than PHO.

Interestingly, our regressions show that Pashto speakers predicted for lower individual *and* social digital capacity. While Pashto and Dari are the two main languages in Afghanistan, language is a sensitive situation [33]. Language is described as a divisive issue in the country, which has led to ethnic conflicts [38]. While Pashto is an official language, there is no standard system for the written language [33], which might complicate literacy rates and language representation in more formalized settings like education and could help explain our results. Occupational data could reveal prior experiences among Pashtuns in non-computing roles as we speculate for PHO; however, more investigation is needed. Our results also show that being a man predicted for lower social digital capacity in ARO. Afghanistan is rated as having a highly masculine culture, where men are observed as assertive and independent, and women are often characterized by patience, tenderness, submissiveness, and affection [35]. Such gender roles are learned early and reinforced in social institutions like schools and work. A deeper investigation is needed to interpret these results fully. It might be worth investigating whether digital literacy is characteristically perceived as feminine among the population and whether men seeking support from others is atypical.

Our results show different responses on the social dimension of ARO in general. Community members' responses to social items might suggest limited experience with social interactions like knowing someone who could teach them any tasks and familiarity with situations like borrowing items or knowing places to go for reliable internet.[5] A closer look might reveal, however, that False responses could signal low individual digital capacity among their social networks (i.e., I don't know someone who can perform these tasks—if I know someone who wanted to help me, they would not be able to do so.). Drawing from Hsiao et al., migrants, which include refugees, "face unique tensions in learning new technologies unavailable or unpopular in their countries of origin," [45, p.135:17] a finding aligned with past CSCW and HCI studies [34, 46, 64].

Meanwhile, social digital literacy items Q26-28 (know someone who would provide me a computer, loan it, or let me come over) relate to physical device asks and overlap with infrastructure (less so questions of expertise). In this case, True responses signal that they know people with such devices and would be open to letting them use them. Such results could reflect Afghan culture as a collectivist and communal culture [35].

While cultural nuances could explain these qualitative results, a deeper investigation of our results is needed. Nevertheless, we extend the large body of work from CSCW and HCI that finds the digital divide is a spectrum influenced by social, cultural, and socio-economic factors, extending Level 3 Digital Divide scholarship beyond individuals and organizations to the community. The social/communal dimension of digital capacity is conceptually meaningful and relates to how supportive a community is, as a whole, for individuals who need digital support. We put forward that strong social digital capacity can, in theory, make up for poor individual digital literacy or help to improve individual digital literacy.

7.1 Limitations and Future Work

Despite being among two distinct populations, we assessed our measure across two cities in a single state within the U.S. Thus, our results are subject to selection bias and need more assessment. As we stated earlier, additional selection bias was introduced by literacy challenges and having women orally respond to surveys, which their husbands entered. This process introduced a layer of second-hand reporting, although ensuring all voices were represented in the survey was important. Nevertheless, our work is exploratory, and represents a first attempt at constructing such a measure.

To address these limitations and to assess whether the measure's structure is similar across demographic groups and locations, the proposed instrument should be implemented across additional communities (including public housing) and in additional states across the U.S. Future efforts should also be made to assess community digital capacity in rural communities. These efforts will enable us to achieve more comprehensive results on community digital capacity and employ confirmatory factor analysis to assess specific hypotheses about the measure's psychometric properties.

Another limitation of our study is the diversity of participant demographics. While we included diverse communities, our sample size and focus on specific underrepresented populations limited further demographic diversification. Future efforts should aim to include a broader range of demographic groups to enrich our understanding of community digital capacity. Additionally, our results suggest that different cultural norms and practices may influence individuals' use of technology and their attitudes toward it. While investigating these cultural factors was outside

[5]See Figure 3 factor analysis. Note responses as "False" to social digital literacy items 20-24 (know_once, know_daily, know_prob, know_teach, know_anytask) and "True" to social digital literacy items 26-28 (know_loan, know_comeover, know_place).

the scope of the current study—where our primary aim was to assess the meaningfulness of our proposed measure—future studies could explore this in more detail. Finally, we plan to engage community members in the evaluation process to ensure the measure is relevant and respectful of the community's needs and values. We aim to continuously iterate the measure based on feedback and changing digital landscapes as we advance.

We anticipate future efforts to address potential cross-loadings of our infrastructure dimension. Because we began with a fairly broad approach in our initial brainstorming of items, it is expected that some items would not fit well. Our infrastructure items did not capture a common theme in a psychometric way that loads across the other two dimensions and might have been cross-loading. In other words, they loaded on social *and* individual dimensions [22][6]. Because cross-loadings complicate our analyses, we removed the items to achieve a clearer factor structure. Nevertheless, the FCC provided national broadband maps to provide infrastructure data, minimizing the need for questions about internet home access. However, they do not cover reliability issues necessary for services raised in prior CSCW and HCI literature like ridesharing, food delivery, telehealthcare, and online education and employment (e.g., [4, 26–28, 31, 48, 51]) that are critical among traditionally underserved communities . Thus, we plan to investigate ways to address potential cross-loadings in our items in the future.

We also removed questions related to digital assets early in our process. These questions were inspired by prior CSCW and HCI research on technology within the context of marginalized communities. The impact of factors related to social media usage (e.g., number of friends or followers, whether people seek online credentialing or obtain online degrees or certificates, and their comfort in requesting help or resources online) on their own might be a worthwhile investigation forward, particularly for CSCW.

8 Conclusion

The proposed Community Digital Capacity Scale is important because it addresses pressing societal issues related to digital inequality and social inclusion. Specifically, it lays the groundwork for more targeted and effective digital literacy interventions in the future. We believe the tool will help assess the role of social ties in providing digital support, especially among underserved communities. Our analyses show that the proposed measure is a quantitatively reliable tool for PHO based on results that confirm internal consistency, and the validity meets the goodness of fit criteria. Qualitatively, our results are promising for ARO but additional work is needed to fully understand the social factor and establish similar levels of reliability as those of PHO.

Acknowledgments

This work was supported by the NSF award CNS-2125012 and the first author completed part of this work as an MIT MLK Visiting Scholar. We thank Alexandra K. Murphy and our Institute of Survey Research colleagues Lisa Holland, Raphael Nishimura, and Wen Chang for providing early insights on this work. We thank our Center for Education Design, Evaluation, and Research colleagues Inah Ko and Kolby Gadd for their early analysis and Shruti Sannon for reviewing an early draft and providing feedback on this work. We thank Diane Cady for editing services and our anonymous reviewers for their supportive feedback. Finally, we thank ChatGPT for helping us to create accessible tables and figures.

[6] I.e., know_provide and social items falling under the question, *"If I lost access to my computer, smartphone, or flip phone/feature phone and needed to use the Internet"* were related to knowing someone who could provide physical access to digital technologies and overlap with infrastructure. Most items of individual digital capacity required internet access (e.g., a_download, a_email, a_videocall, a_purchase).

Acknowledgments

This work was supported by the NSF award CNS-2125012 and the first author completed part of this work as an MIT MLK Visiting Scholar. We extend our deepest appreciation to our community partners and members for their invaluable support and participation. We thank Alexandra K. Murphy and our Institute of Survey Research colleagues Lisa Holland, Raphael Nishimura, and Wen Chang for providing early insights. We thank our Center for Education Design, Evaluation, and Research colleagues Inah Ko and Kolby Gadd for their early analysis and Shruti Sannon for reviewing an early draft and providing feedback on this work. We thank Diane Cady for editing services and our anonymous reviewers for their supportive feedback. Finally, we thank ChatGPT for helping us to create accessible tables and figures.

References

[1] [n. d.]. WhatsApp to Reach and Engage RIM Communities | National Resource Center for Refugees, Immigrants, and Migrants (NRC-RIM) — nrcrim.org. https://nrcrim.org/whatsapp-reach-and-engage-rim-communities. [Accessed 05-01-2024].

[2] Carol E Adair, Anna C Holland, Michelle L Patterson, Kate S Mason, Paula N Goering, Stephen W Hwang, and At Home/Chez Soi Project Team. 2012. Cognitive interviewing methods for questionnaire pre-testing in homeless persons with mental disorders. Journal of Urban Health 89 (2012), 36–52.

[3] Najeh Aissaoui. 2021. The digital divide: A literature review and some directions for future research in light of COVID-19. Global Knowledge, Memory and Communication, Ahead-of-Print. Online ahead-of-print (2021).

[4] Marcy G Antonio, Alicia Williamson, Vaishnav Kameswaran, Ashley Beals, Elizabeth Ankrah, Shannon Goulet, Yucen Wang, Grecia Macias, Jade James-Gist, Lindsay K Brown, Sage Davis, Srijanani Pillai, Lorraine Buis, Tawanna Dillahunt, and Tiffany C Veinot. 2023. Targeting patients' cognitive load for telehealth video visits through student-delivered helping sessions at a United States federally qualified health center: equity-focused, mixed methods pilot intervention study. Journal of Medical Internet Research 25 (2023), e42586.

[5] Anteneh Ayanso, Danny I Cho, and Kaveepan Lertwachara. 2011. ICT development index and the digital divide. (2011).

[6] Amna Batool, Tawanna R Dillahunt, Julie Hui, and Mustafa Naseem. 2024. Gendered, Collectivist Journeys: Exploring Sociotechnical Adaptation Among Afghan Refugees in the USA. Proc. ACM Hum.-Comput. Interact. 8, CSCW2, Article 498 (Nov. 2024), 32 pages. doi:10.1145/3687037

[7] Fran Baum, Lareen Newman, and Katherine Biedrzycki. 2012. Vicious cycles: digital technologies and determinants of health in Australia. Health promotion international 29, 2 (2012), 349–360.

[8] Paul C Beatty and Gordon B Willis. 2007. Research synthesis: The practice of cognitive interviewing. Public opinion quarterly 71, 2 (2007), 287–311.

[9] Elisabeth Beaunoyer, Sophie Dupéré, and Matthieu J Guitton. 2020. COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. Computers in human behavior 111 (2020), 106424.

[10] Jacob Belliveau and Igor Yakovenko. 2022. Evaluating and improving the quality of survey data from panel and crowd-sourced samples: A practical guide for psychological research. Experimental and Clinical Psychopharmacology 30, 4 (2022), 400.

[11] Silvia Cristina Bettez and Kathy Hytten. 2013. Community building in social justice work: A critical approach. Educational Studies 49, 1 (2013), 45–66.

[12] Godfred O Boateng, Torsten B Neilands, Edward A Frongillo, Hugo R Melgar-Quiñonez, and Sera L Young. 2018. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Frontiers in public health 6 (2018), 149.

[13] Walter R Boot, Neil Charness, Sara J Czaja, Joseph Sharit, Wendy A Rogers, Arthur D Fisk, Tracy Mitzner, Chin Chin Lee, and Sankaran Nair. 2015. Computer proficiency questionnaire: assessing low and high computer proficient seniors. The Gerontologist 55, 3 (2015), 404–411.

[14] Timothy A Brown. 2015. Confirmatory factor analysis for applied research. Guilford publications.

[15] Timothy A Brown and Michael T Moore. 2012. Confirmatory factor analysis. Handbook of structural equation modeling 361 (2012), 379.

[16] Eunseong Cho and Seonghoon Kim. 2015. Cronbach's coefficient alpha: Well known but poorly understood. Organizational research methods 18, 2 (2015), 207–230.

[17] Virginia Cobigo, Lynn Martin, and Rawad Mcheimech. 2016. Understanding community. Canadian Journal of Disability Studies 5, 4 (2016), 181–203.

[18] Sheldon Cohen and Harry M Hoberman. 1983. Positive events and social supports as buffers of life change stress. Journal of applied social psychology 13, 2 (1983), 99–125.

[19] James S Coleman. 1988. Social capital in the creation of human capital. American journal of sociology 94 (1988), S95–S120.

[20] Jessica Colnago, Lorrie Faith Cranor, Alessandro Acquisti, and Kate Hazel Stanton. 2022. Is it a concern or a preference? An investigation into the ability of privacy scales to capture and distinguish granular privacy constructs. In Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022). 331–346.

[21] Patricia Costa, Jonatan Castano-Munoz, and Panagiotis Kampylis. 2021. Capturing schools' digital capacity: Psychometric analyses of the SELFIE self-reflection tool. Computers & Education 162 (2021), 104080.

[22] Anna B Costello and Jason Osborne. 2019. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical assessment, research, and evaluation 10, 1 (2019), 7.

[23] Brian Detlor, Heidi Julien, Tara La Rose, and Alexander Serenko. 2022. Community-led digital literacy training: Toward a conceptual framework. Journal of the Association for Information Science and Technology 73, 10 (2022), 1387–1400.

[24] Sanjeev Dewan and Frederick J Riggins. 2005. The digital divide: Current and future research directions. Journal of the Association for information systems 6, 12 (2005), 298–337.

[25] Bidit Dey, Karim Sorour, and Raffaele Filieri. 2016. ICTs in developing countries: research, practices and policy implications. Springer.

[26] Tawanna R. Dillahunt, Vaishnav Kameswaran, Linfeng Li, and Tanya Rosenblat. 2017. Uncovering the Values and Constraints of Real-Time Ridesharing for Low-Resource Populations. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 2757–2769. doi:10.1145/3025453.3025470

[27] Tawanna R Dillahunt, Vaishnav Kameswaran, Desiree McLain, Minnie Lester, Delores Orr, and Kentaro Toyama. 2018. Entrepreneurship and the socio-technical chasm in a lean economy. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems CHI '18 (2018), Paper 240, 14 pages.

[28] Tawanna R. Dillahunt, Jason Lam, Alex Lu, and Earnest Wheeler. 2018. Designing Future Employment Applications for Underserved Job Seekers: A Speed Dating Study. In Proceedings of the 2018 Designing Interactive Systems Conference (Hong Kong, China) (DIS '18). Association for Computing Machinery, New York, NY, USA, 33–44. doi:10.1145/3196709.3196770

[29] Tawanna R Dillahunt, Alex Jiahong Lu, Aarti Israni, Ruchita Lodha, Savana Brewer, Tiera S Robinson, Angela Brown Wilson, and Earnest Wheeler. 2022. The Village: Infrastructuring Community-based Mentoring to Support Adults Experiencing Poverty. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 1–17.

[30] Tawanna R Dillahunt, Kerby Shedden, Mila Ekaterina Filipof, Soyoung Lee, Mustafa Naseem, Kentaro Toyama, and Julie Hui. 2024. Toward a Measure of Collective Digital Capacity: An Exploratory Analysis. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI EA '24). Association for Computing Machinery, New York, NY, USA, Article 329, 8 pages. doi:10.1145/3613905.3650811

[31] Tawanna R. Dillahunt, Sylvia Simioni, and Xuecong Xu. 2019. Online Grocery Delivery Services: An Opportunity to Address Food Disparities in Transportation-Scarce Areas. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3290605.3300879

[32] Paul DiMaggio, Eszter Hargittai, Coral Celeste, Steven Shafer, et al. 2004. From unequal access to differentiated use: A literature review and agenda for research on digital inequality. Social inequality 1 (2004), 355–400.

[33] Muhammad Ali Dinakhel. 2018. An Analysis of Conflict between Pashto and Dari Languages of Afghanistan. Central Asia 83, Winter (2018), 79–99.

[34] Bryan Dosono and Bryan Semaan. 2019. Moderation practices as emotional labor in sustaining online communities: The case of AAPI identity work on Reddit. In Proceedings of the 2019 CHI conference on human factors in computing systems. 1–13.

[35] Ehsan M Entezar. 2008. Afghanistan 101: understanding Afghan culture. Xlibris Corporation.

[36] Emily Fisher. 2020. Connect 313 Works to Bridge the Digital Divide in Detroit. Retrieved December 30, 2023 from https://detroitisit.com/connect-313-bridge-the-digital-divide-in-detroit/

[37] Rikin Gandhi, Rajesh Veeraraghavan, Kentaro Toyama, and Vanaja Ramprasad. 2007. Digital green: Participatory video for agricultural extension. In 2007 International conference on information and communication technologies and development. IEEE, 1–10.

[38] Saran Kaur Gill. 2013. Language policy challenges in multi-ethnic Malaysia. Vol. 8. Springer Science & Business Media.

[39] Amy L Gonzales, Lindsay Ems, and Venkata Ratnadeep Suri. 2016. Cell phone disconnection disrupts access to healthcare and health resources: A technology maintenance perspective. new media & society 18, 8 (2016), 1422–1438.

[40] Elisa J Gordon. 2000. When oral consent will do. Field Methods 12, 3 (2000), 235–238.

[41] Louis Guttman. 1945. A basis for analyzing test-retest reliability. Psychometrika 10, 4 (1945), 255–282.

[42] Martin S Hagger, Daniel F Gucciardi, and Nikos LD Chatzisarantis. 2017. On nomological validity and auxiliary assumptions: The importance of simultaneously testing effects in social cognitive theories applied to health behavior and some guidelines. Frontiers in psychology 8 (2017), 1933.

[43] Eszter Hargittai. 2001. Second-level digital divide: Mapping differences in people's online skills. arXiv preprint cs/0109068 (2001).

[44] Rakibul Hasan, Rebecca Weil, Rudolf Siegel, and Katharina Krombholz. 2023. A Psychometric Scale to Measure Individuals' Value of Other People's Privacy (VOPP). In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 1–14.

[45] Joey Chiao-Yin Hsiao, Sylvia Darling, and Tawanna R. Dillahunt. 2023. How Recent Migrants Develop Trust Through Community Commerce: The Emergence of Sociotechnical Adaptation. Proc. ACM Hum.-Comput. Interact. 7, CSCW1, Article 135 (apr 2023), 24 pages. doi:10.1145/3579611

[46] Joey Chiao-Yin Hsiao and Tawanna R Dillahunt. 2018. Technology to support immigrant access to social capital and adaptation to a new country. Proceedings of the ACM on Human-Computer Interaction 2, CSCW (2018), 1–21.

[47] Li-tze Hu, Peter M Bentler, et al. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling 6, 1 (1999), 1–55.

[48] Julie Hui, Nefer Ra Barber, Wendy Casey, Suzanne Cleage, Danny C Dolley, Frances Worthy, Kentaro Toyama, and Tawanna R Dillahunt. 2020. Community collectives: Low-tech social support for digitally-engaged entrepreneurship. In Proceedings of the 2020 CHI conference on human factors in computing systems. 1–15.

[49] Julie Hui, Kristin Seefeldt, Christie Baer, Lutalo Sanifu, Aaron Jackson, and Tawanna R Dillahunt. 2023. Community Tech Workers: Scaffolding Digital Engagement Among Underserved Minority Businesses. Proceedings of the ACM on Human-Computer Interaction 7, CSCW2 (2023), 1–25.

[50] Julie Hui, Kristin Seefeldt, Lutalo Sanifu, Christie Baer, Jeanette Szomstein, and Tawanna R Dillahunt. 2024. "I was able to give her the confidence": Reciprocal Capacity Building in a Community-based Program for Digital Engagement. In Proceedings of the CHI Conference on Human Factors in Computing Systems. 1–13.

[51] Julie Hui, Kentaro Toyama, Joyojeet Pal, and Tawanna Dillahunt. 2018. Making a living my way: Necessity-driven entrepreneurship in resource-constrained communities. Proceedings of the ACM on Human-Computer Interaction 2, CSCW (2018), 1–24.

[52] Amy E Hurley, Terri A Scandura, Chester A Schriesheim, Michael T Brannick, Anson Seers, Robert J Vandenberg, and Larry J Williams. 1997. Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. Journal of organizational behavior (1997), 667–683.

[53] Immigration Research Initiative. [n. d.]. Refugee Resettlement per Capita: Which States Do the Most? - Immigration Research Initiative — immresearch.org. https://immresearch.org/publications/refugee-resettlement-per-capita-which-states-do-the-most/. [Accessed 21-06-2024].

[54] Karl G Jöreskog. 1969. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika 34, 2 (1969), 183–202.

[55] Siriwatchana Kaeophanuek, Jaitip Na-Songkhla, and Prachyanun Nilsook. 2018. How to enhance digital literacy skills among. International Journal of Information and Education Technology 8, 4 (2018), 292–297.

[56] Andrea Karpati. 2011. Digital Literacy in Education. http://unesdoc.unesco.org/images/0021/002144/214485e.pdf. United Nations Educational, Scientific, and Cultural Organization (UNESCO) Institute for Information Technologies in Education (2011). [Accessed 18-12-2023].

[57] Naveena Karusala, Azra Ismail, Karthik S Bhat, Aakash Gautam, Sachin R Pendse, Neha Kumar, Richard Anderson, Madeline Balaam, Shaowen Bardzell, Nicola J Bidwell, Melissa Densmore, Elizabeth Kaziunas, Anne Marie Piper, Noopur Raval, Pushpendra Singh, Austin Toombs, Nervo Verdezoto, and Ding Wang. 2021. The Future of Care Work: Towards a Radical Politics of Care in CSCW Research and Practice. In Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing (Virtual Event, USA) (CSCW '21 Companion). Association for Computing Machinery, New York, NY, USA, 338–342. doi:10.1145/3462204.3481734

[58] Clare Killen, Helen Beetham, and Sarah Knight. 2017. Developing organisational approaches to digital capability. JISC Guides. JISC. https://www. jisc. ac. uk/guides/developing-organisationalapproaches-to-digital-capability (2017).

[59] Paul Kline. 2013. Handbook of psychological testing. Routledge.

[60] Yasmine Kotturi, Julie Hui, TJ Johnson, Lutalo Sanifu, and Tawanna R Dillahunt. 2024. Sustaining Community-Based Research in Computing: Lessons from Two Tech Capacity Building Initiatives for Local Businesses. Proceedings of the ACM on Human-Computer Interaction 8, CSCW1 (2024), 1–31.

[61] Sandjar Kozubaev, Fernando Rochaix, Carl DiSalvo, and Christopher A Le Dantec. 2019. Spaces and traces: Implications of smart technology in public housing. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 1–13.

[62] Soyoung Lee, Julie Hui, and Tawanna R Dillahunt. 2025. Sociocultural Factors in Digital Skills Learning: A Community-Based Intervention Among U.S. Public Housing Adults. Proc. ACM Hum.-Comput. Interact. 8, CSCW2 (2025). in press.

[63] Soyoung Lee, Julie Hui, Zachary Rowe, and Tawanna R Dillahunt. 2023. A Collective Approach to Providing Digital Skills Training Among U.S. Public Housing Residents. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 4, 6 pages. doi:10.1145/3544549.3585712

[64] Jessica Lingel, Mor Naaman, and Danah M Boyd. 2014. City, self, network: transnational migrants and online identity work. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. 1502–1510.

[65] Sonia Livingstone and Ellen Helsper. 2007. Gradations in digital inclusion: Children, young people and the digital divide. New media & society 9, 4 (2007), 671–696.

[66] Sophie Lythreatis, Sanjay Kumar Singh, and Abdul-Nasser El-Kassar. 2022. The digital divide: A review and future research agenda. Technological Forecasting and Social Change 175 (2022), 121359.

[67] Richard G Netemeyer, William O Bearden, and Subhash Sharma. 2003. Scaling procedures: Issues and applications. sage publications.

[68] Eli Newman. 2021. The Digital Divide in Detroit: How the City is Working to Bridge the Gaps. Retrieved December 30, 2023 from https://wdet.org/2021/06/28/the-digital-divide-in-detroit-how-the-city-is-working-to-bridge-the-gaps/

[69] Alex Pentland, Richard Fletcher, and Amir Hasson. 2004. Daknet: Rethinking connectivity in developing nations. Computer 37, 1 (2004), 78–83.

[70] Public and Affordable Housing Research Corporation (PAHRC). 2018. EMPLOYMENT TRENDS AMONG PEOPLE LIVING IN PUBLICLY SUPPORTED HOMES. https://www.pahrc.org/wp-content/uploads/2019/10/Employment-Trends-Among-People-Living-in-Publicly-Supported-Homes.pdf. [Accessed 11-01-2024].

[71] Robert D Putnam. 2000. Bowling alone: The collapse and revival of American community. Simon and schuster.

[72] Massimo Ragnedda and Glenn Muschert. 2016. Theorizing digital divides and digital inequalities. (2016).

[73] Divya Ramachandran, John Canny, Prabhu Dutta Das, and Edward Cutrell. 2010. Mobile-izing health workers in rural India. In Proceedings of the SIGCHI conference on human factors in computing systems. 1889–1898.

[74] Tenko Raykov and George A Marcoulides. 2011. Introduction to psychometric theory. Routledge.

[75] Glenna L Read, Harry Yaojun Yan, Philip B Anderson, Laura PB Partain, Zachary Vaughn, Antonina Semivolos, Yeweon Kim, and Amy L Gonzales. 2022. Making stability dependable: stable cellphone access leads to better health outcomes for those experiencing poverty. Information, Communication & Society 25, 14 (2022), 2122–2139.

[76] Laura Robinson, Jeremy Schulz, Grant Blank, Massimo Ragnedda, Hiroshi Ono, Bernie Hogan, Gustavo S Mesch, Shelia R Cotten, Susan B Kretchmer, Timothy M Hale, et al. 2020. Digital inequalities 2.0: Legacy inequalities in the information age. First Monday 25, 7 (2020).

[77] Nithya Sambasivan and Thomas Smyth. 2010. The Human Infrastructure of ICTD. In Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development (London, United Kingdom) (ICTD '10). Association for Computing Machinery, New York, NY, USA, Article 40, 9 pages. doi:10.1145/2369220.2369258

[78] Anique Scheerder, Alexander Van Deursen, and Jan Van Dijk. 2017. Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. Telematics and informatics 34, 8 (2017), 1607–1624.

[79] John A Schinka and Wayne F Velicer. 2003. Handbook of Psychology, Developmental Psychology. Vol. 6. John Wiley & Sons.

[80] Klaas Sijtsma. 2009. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika 74 (2009), 107–120.

[81] Maria Spante, Sylvana Sofkova Hashemi, Mona Lundin, and Anne Algers. 2018. Digital competence and digital literacy in higher education research: Systematic review of concept use. Cogent Education 5, 1 (2018), 1519143.

[82] Hiller Spires and Melissa Bartlett. 2012. Digital literacies and learning: Designing a path forward. Friday Institute White Paper Series 5 (2012), 1–24.

[83] Holger Steinmetz. 2019. Is Cronbach's alpha necessary when a confirmatory factor analysis has been done?

[84] Michael Stern, Alison Adams, and Shaun Elsasser. 2009. How levels of Internet proficiency affect usefulness of access across rural, suburban, and urban communities. Sociological Inquiry 79, 4 (2009), 391–417.

[85] Rodney L Terry and Aleia Clark Fobia. 2017. Cognitive Pretesting of the 2015 CPS Computer and Internet Use Supplement. Survey Methodology (2017), 07.

[86] Alexander Van Deursen and Jan Van Dijk. 2011. Internet skills and the digital divide. New media & society 13, 6 (2011), 893–911.

[87] Alexander van Deursen, Jan van Dijk, and Ellen Helsper. 2014. Investigating outcomes of online engagement. (2014).

[88] Alexander J.A.M. Van Deursen and Ellen J Helsper. 2015. The third-level digital divide: Who benefits most from being online? In Communication and information technologies annual. Emerald Group Publishing Limited, 29–52.
[89] Alexander J.A.M. Van Deursen, Ellen J Helsper, and Rebecca Eynon. 2016. Development and validation of the Internet Skills Scale (ISS). Information, communication & society 19, 6 (2016), 804–823.
[90] Alexander JAM Van Deursen and Jan AGM Van Dijk. 2014. The digital divide shifts to differences in usage. New media & society 16, 3 (2014), 507–526.
[91] Alexander J.A.M. Van Deursen and Jan AGM Van Dijk. 2019. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. New media & society 21, 2 (2019), 354–375.
[92] Jan AGM Van Dijk. 2006. Digital divide research, achievements and shortcomings. Poetics 34, 4-5 (2006), 221–235.
[93] Johannes AGM van Dijk and Kenneth Hacker. 2003. The digital divide as a complex and dynamic phenomenon. The information society 19, 4 (2003), 315–326.
[94] Mark Warschauer. 2004. Technology and social inclusion: Rethinking the digital divide. MIT press.
[95] Dmitri Williams. 2006. On and off the'Net: Scales for social capital in an online era. Journal of computer-mediated communication 11, 2 (2006), 593–628.
[96] Xuecong Xu, Xiang Yan, and Tawanna R. Dillahunt. 2019. Reaching Hard-To-Reach Populations: An Analysis of Survey Recruitment Methods. In Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing (Austin, TX, USA) (CSCW '19 Companion). Association for Computing Machinery, New York, NY, USA, 428–432. doi:10.1145/3311957.3359447

A Cohort Means

For simplicity, Table codes were generated with Chat GPT 4.0 and modified slightly by the authors.
Table 3 includes a list of item means across both cohorts.

No.	Item	PHO Means	ARO Means
1	a_text	2.836	2.389
2	a_keyboard	2.818	2.463
3	a_download	2.764	2.358
4	a_watch	2.855	2.568
5	a_record	2.636	2.474
6	a_openread	2.764	1.716
7	a_email	2.455	1.316
8	a_search	2.709	1.916
9	a_calendar	2.345	1.642
10	a_videocall	2.655	2.347
11	a_payment	2.655	1.558
12	a_purchase	2.655	1.484
13	a_socialmedia	2.655	2.642
14	a_prvsetting	2.527	2.021
15	a_spreadsheet	2.309	1.158
16	i_address	1.855	1.905
17	i_reliable	1.855	1.916
18	i_place	1.836	1.863
19	i_outside	1.927	1.958
20	k_once	1.945	1.937
21	k_daily	1.782	1.905
22	k_prob	2.418	2.684
23	k_teach	1.800	1.821
24	k_anytask	1.655	1.726
25	k_provide	1.691	1.768
26	k_loan	1.673	1.179
27	k_comeover	1.782	1.126
28	k_place	1.764	1.084

Table 3. Comparison of Item Means across PHO and ARO

B Participant Demographics

See Tables 4 and 5 (Table 5 for ARO only).

		Public Housing (n=56)	**Afghan Refugee Org. (n=97)**
Age	10s	1 (2%)	1(1%)
	20s	5 (9%)	45 (46%)
	30s	10 (18%)	23 (24%)
	40s	15 (27%)	13 (13%)
	50s	17 (30%)	4 (4%)
	60s	4 (7%)	–
	70s	1 (2%)	–
		53 (95%)	**87 (90%)**
Gender	Male	16 (29%)	57 (59%)
	Female	35 (63%)	38 (39%)
	Non-binary	1 (2%)	–
	Self-describe	1 (2%)	–
		53 (95%)	**95 (98%)**
Race/ Ethnicity	White	16 (29%)	–
	Black	40 (71%)	–
	Hispanic Latino, or Spanish origin	–	1 (1%)
	American Indian	1 (2%)	–
	Native Hawaiian/ Pacific Islander	2 (4%)	–
	Asian	–	92 (95%)
	Some other race, ethnicity or origin	–	4 (4%)
		59 (105%)	**97 (100%)**

Table 4. Respondent Demographics. Note that participants could enter more than one race.

		All (n=97)	**Pashto (n=77)**	**Dari (n=20)**
Age	10s	1 (1%)	1(1%)	–
	20s	45 (46%)	35 (45%)	10 (50%)
	30s	23 (24%)	15 (19%)	8 (40%)
	40s	13 (13%)	12 (16%)	1(5%)
	50s	4 (4%)	3 (4%)	1 (5%)
	60s	– (–%)	–	–
	70s	– (–)	–	–
		87 (90%)	**66 (86%)**	**20 (100%)**
Gender	Male	57 (59%)	45 (58%)	12 (60%)
	Female	38 (39%)	30 (39%)	8 (40%)
	Non-binary	–	–	–
	Self-describe	–	–	–
		95 (98%)	**75 (97%)**	**20 (100%)**
Race/ Ethnicity	White	–		
	Black	–	–	–
	Hispanic Latino, or Spanish origin	1 (1%)	–	1 (5%)
	American Indian	–	–	–
	Native Hawaiian/ Pacific Islander	–	–	–
	Asian	92 (95%)	74 (96%)	18 (90%)
	Some other race, ethnicity or origin	4 (4%)	4 (5%)	–
		97 (100%)	**78 (101%)**	**19 (95%)**

Table 5. Demographic Breakdown for Pashto and Dari for ARO only. Note that participants could enter more than one race.

C Exploratory Factor Analysis (EFA) Factor Loadings

This section presents the factor loadings derived from both Exploratory EFA. In EFA, the factor loadings indicate how strongly each variable is associated with the potential underlying factors identified in the analysis, helping to uncover the latent structure of the data without any prior assumptions. In contrast, CFA involves testing a hypothesized structure by specifying which variables are expected to load on each factor beforehand. The factor loadings in CFA measure how well the data fits this predetermined structure, confirming or refuting the proposed model. Table 6 includes the PHO and ARO EFA Factor Loadings.

D Regression Analyses

D.1 PHO CFA Regression: Factor 1

Please refer to Tables 7 and 8.

D.2 PHO CFA Regression: Factor 2

Please refer to Tables 9 and 10.

	PHO EFA				ARO EFA			
Variable	**f1**	**f2**	**srmr**	**uniqueness**	**f1**	**f2**	**srmr**	**uniqueness**
a_text	0.7220	0.0584	0.1069	0.4752	0.8285	0.0000	0.0660	0.3136
a_keyboard	0.5563	0.0891	0.0770	0.6825	0.8702	0.0000	0.0761	0.2427
a_download	0.7916	-0.1402	0.0721	0.3537	0.8482	0.0000	0.0716	0.2805
a_watch	0.5838	-0.2083	0.1013	0.6158	0.8598	0.0000	0.1039	0.2607
a_record	0.5782	-0.0059	0.0914	0.6657	0.8322	0.0000	0.1045	0.3075
a_openread	0.7362	0.0434	0.0884	0.4561	0.6108	0.0000	0.1872	0.6269
a_email	0.7389	-0.0132	0.0862	0.4538	0.4432	0.0000	0.2779	0.8036
a_search	0.6877	-0.0321	0.0858	0.5261	0.7796	0.0000	0.1450	0.3923
a_calendar	0.5885	-0.1406	0.1027	0.6340	0.6370	0.0000	0.1707	0.5943
a_videocall	0.6397	-0.0790	0.0938	0.5845	0.6943	0.0000	0.0905	0.5180
a_payment	0.7489	-0.1848	0.0724	0.4050	0.5493	0.0000	0.2210	0.6983
a_purchase	0.7676	-0.0666	0.0936	0.4064	0.5594	0.0000	0.2634	0.6871
a_socialmedia	0.6894	0.1662	0.0974	0.4971	0.5857	0.0000	0.0639	0.6569
a_prvsetting	0.6325	0.2481	0.0986	0.5384	0.7851	0.0000	0.1200	0.3836
a_spreadsheet	0.5071	0.1828	0.1167	0.7095	–	–	–	–
i_address	0.2425	-0.0472	0.1496	0.9389	–	–	–	–
i_reliable	0.2353	0.1289	0.1277	0.9280	–	–	–	–
i_place	0.2769	0.0504	0.1449	0.9208	–	–	–	–
i_outside	0.3438	0.2219	0.1220	0.8326	–	–	–	–
k_once	-0.0659	0.3542	0.0794	0.8702	0.0000	0.1940	0.1499	0.9624
k_daily	-0.1551	0.5599	0.0829	0.6625	–	–	–	–
k_prob	-0.1910	0.4079	0.0657	0.7971	–	–	–	–
k_teach	0.1573	0.6917	0.0666	0.4968	0.0000	0.8017	0.1175	0.3572
k_anytask	0.0535	0.5718	0.0887	0.6702	0.0000	0.8013	0.1298	0.3579
k_provide	0.0361	0.6405	0.0721	0.5885	0.0000	0.6196	0.1097	0.6161
k_loan	0.0222	0.5399	0.1032	0.7080	0.0000	-0.2058	0.2422	0.9576
k_comeover	0.0916	0.4984	0.0849	0.7432	0.0000	-0.2346	0.2895	0.9449
k_place	0.2988	0.3368	0.1014	0.7973	–	–	–	–

Table 6. Factor Loadings for PHO and ARO: Overall EFA SRMR for PHO=0.0961 and ARO=0.1621

D.3 ARO CFA Regression: Factor 1

Please refer to Tables 11 and 12.

D.4 ARO CFA Regression: Factor 2

Please refer to Tables 13 and 14.

Received January 2024; revised July 2024; accepted October 2024

Variable	β	p	95% CI (LL, UL)
Intercept	24.65	< .001	[19.01, 30.29]
Sex (Male)	0.84	.558	[-2.04, 3.72]
Sex (Nonbinary)	3.40	.443	[-5.46, 12.27]
Sex (Self-describe)	3.33	.454	[-5.55, 12.20]
Age	0.01	.912	[-0.11, 0.12]

Table 7. Abridged PHO CFA Regression results for Factor 1.

Variable	β	p	95% CI (LL, UL)
Intercept	22.79	.001	[10.24, 35.34]
Sex (Male)	0.77	.652	[-2.68, 4.22]
Sex (Nonbinary)	0.28	.946	[-7.97, 8.53]
Sex (Self-describe)	4.14	.341	[-4.60, 12.88]
Age	-0.03	.677	[-0.20, 0.13]
HHS	-0.29	.648	[-1.60, 1.01]
Education	1.61	.030	[0.16, 3.06]
Work	-4.41	.018	[-8.00, -0.83]
Money	0.05	.949	[-1.56, 1.66]

Table 8. Abridged PHO CFA Regression results for Factor 1.

Variable	β	p	95% CI (LL, UL)
Intercept	6.13	< .001	[4.85, 7.41]
Sex (Male)	0.16	.631	[-0.50, 0.81]
Sex (Nonbinary)	0.89	.379	[-1.13, 2.91]
Sex (Self-describe)	-1.45	.156	[-3.47, 0.57]
Age	-0.01	.480	[-0.04, 0.02]

Table 9. Abridged PHO CFA Regression results for Factor 2

Variable	β	p	95% CI (LL, UL)
Intercept	6.10	.002	[2.35, 9.85]
Sex (Male)	0.29	.567	[-0.74, 1.32]
Sex (Nonbinary)	0.53	.664	[-1.93, 2.99]
Sex (Self-describe)	-1.60	.219	[-4.21, 1.01]
Age	-0.02	.470	[-0.07, 0.03]
HHS	0.09	.644	[-0.30, 0.48]
Education	0.07	.755	[-0.37, 0.50]
Work	-0.14	.790	[-1.21, 0.93]
Money	-0.06	.791	[-0.54, 0.42]

Table 10. Abridged PHO CFA Regression results for Factor 2.

Variable	β	p	95% CI (LL, UL)
Intercept	26.04	< .001	[20.24, 31.84]
Sex (Male)	1.01	.499	[-1.95, 3.96]
User Language (Pashto)	-3.52	.040	[-6.88, -0.16]
Age	-0.09	.291	[-0.26, 0.08]

Table 11. Abridged ARO CFA Regression results for Factor 1.

Variable	β	p	95% CI (LL, UL)
Intercept	26.85	< .001	[19.41, 34.28]
Sex (Male)	0.14	.933	[-3.16, 3.44]
User Language (Pashto)	-3.11	.090	[-6.73, 0.50]
Age	-0.05	.580	[-0.24, 0.13]
HHS	-0.39	.132	[-0.91, 0.12]
Education	0.10	.882	[-1.27, 1.47]

Table 12. Abridged ARO Regression results for Factor 1.

Variable	β	p	95% CI (LL, UL)
Intercept	3.49	< .001	[2.75, 4.23]
Sex (Male)	-0.44	.021	[-0.82, -0.07]
User Language (Pashto)	-0.40	.069	[-0.82, 0.03]
Age	0.03	.009	[0.01, 0.05]

Table 13. Abridged ARO CFA Regression results for Factor 2.

Variable	β	p	95% CI (LL, UL)
Intercept	3.86	< .001	[2.92, 4.81]
Sex (Male)	-0.31	.144	[-0.73, 0.11]
User Language (Pashto)	-0.51	.031	[-0.96, -0.05]
Age	0.02	.058	[-0.00, 0.05]
HHS	0.02	.509	[-0.04, 0.09]
Education	-0.14	.113	[-0.31, 0.03]

Table 14. Abridged ARO Regression results for Factor 2.