Development Engineering 8 (2023) 100107

Contents lists available at ScienceDirect

Development Engineering

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

Validated digital literacy measures for populations with low levels of internet experiences

Ayesha Ali [a,*], Agha Ali Raza [b], Ihsan Ayyub Qazi [b]

[a] Department of Economics, Lahore University of Management Sciences, DHA, Lahore Cantt, 54792, Lahore, Pakistan
[b] Department of Computer Science, Lahore University of Management Sciences, DHA, Lahore Cantt, 54792, Lahore, Pakistan

ARTICLE INFO

JEL classification:
C83
D83
L86
L82
O10

Keywords:
Digital literacy
Survey measures
Participant observation
Validation
Machine learning

ABSTRACT

A growing body of evidence suggests that digital literacy is an important barrier constraining adoption and use of Internet and digital technologies in the developing world. By enabling people to effectively find valuable information online, digital literacy can play a crucial role in expanding economic opportunities, thereby leading to human development and poverty reduction. Unfortunately, there is a dearth of validated survey measures for capturing digital literacy of populations who have limited prior exposure to technology. We present a novel approach for measuring digital literacy of low literacy and new Internet users, an important segment of users in developing countries. Using a sample of 143 social media users in Pakistan, which includes a significant fraction of low literacy individuals, we measure digital literacy by observing the effectiveness of participants in completing a series of tasks and by recording a set of self-reported survey responses. We then use machine learning methods (e.g., Random Forest) to identify a parsimonious set of survey questions that are most predictive of ground truth digital literacy established through participant observation. Our approach is easily scalable in low-resource settings and can aid in tracking digital literacy as well as designing interventions and policies tailored to users with different levels of digital literacy.

1. Introduction

Digital technologies can play an important role in alleviating poverty and inequality by enabling access to economic opportunities (Jack and Suri, 2014; Chun and Tang, 2018; Hjort and Tian, 2021). However, 37% of the world's population or nearly 2.9 billion people remain offline, 96% of whom live in developing countries. This is despite the fact that 94% of the population in developing countries is covered by at least a 3G cellular network.[1] One major barrier to closing the digital divide is the lack of digital literacy (Dimaggio et al. 2004, Zillien and Hargittai 2009, Rains and Tsetsi 2017, Hargittai and Micheli 2019). Digital literacy—defined as the ability to access and effectively find information online (Hargittai, 2005; Gilster, 1997) is the most often cited reason for why individuals are held back from taking up the Internet (World Bank, 2021).[2]

Digital literacy matters not only for Internet adoption, but also for effectively finding information in the digital space. Evidence from developing countries shows that digital technologies provide access to valuable information about markets, jobs, health, educational and financial services, but their benefits depend on complementary investments that enable effective use of these technologies such as infrastructure and skills (Aker and Blumenstock 2014, Wheeler et al. 2022, Dodson et al. 2013). Thus, digital literacy by enabling effective use of the digital technologies, can play a crucial role in expanding economic opportunities, thereby leading to human development and poverty reduction. Furthermore, several studies show that individuals with higher digital literacy are better at spotting fake news and misleading content online (Ali and Qazi 2022; Sirlin et al. 2021; Muda et al. 2021; Flintham et al. 2018). Thus, digital literacy can help individuals become more discerning consumers of online content, which can in turn have positive effects on social and political behaviors (Guriev et al. 2020; Levy 2021; Zhuravskaya et al. 2020; Iyengar et al. 2019).

Despite the importance of digital literacy from a development perspective, there is a dearth of validated survey measures to capture digital literacy especially for new Internet users and populations with low levels of literacy. Prior works on measuring digital literacy have focused on developed countries, where age has been found to be a key moderator for digital literacy (Grinberg et al. 2019, Hargittai et al.

* Corresponding author.
E-mail addresses: ayeshaali@lums.edu.pk (A. Ali), agha.ali.raza@lums.edu.pk (A.A. Raza), ihsan.qazi@lums.edu.pk (I.A. Qazi).

[1] Data from International Telecommunications Union 2021 global and regional ICT statistics (ITU, 2021).

[2] A World Bank survey of 22 developing countries in 2017-18 found that 69% of the respondents cited lack of digital literacy as the reason for not taking up data services. Responses included "*Do not know what internet is*" and "*Do not know how to use internet*" (World Bank, 2021).

https://doi.org/10.1016/j.deveng.2023.100107
Received 24 June 2022; Received in revised form 27 December 2022; Accepted 2 January 2023
Available online 7 January 2023

2018, Guess and Munger 2020, Brashier and Schacter 2020). However, in developing countries a larger set of demographic attributes are likely to be correlated with digital literacy (e.g., education, gender, and income level) due to the unique barriers faced when using the Internet (Medhi Thies 2015, Vashistha et al. 2019, Qazi et al. 2021). Self-reported survey measures of digital literacy can serve as a proxy but are likely to be contaminated with social desirability bias or random responses (Osborne and Blanchard, 2011; Antin and Shaw, 2012). Objective measurement of digital literacy would involve observing individuals complete a set of digital literacy tasks, resulting in verifiable measurements (e.g., proportion of tasks successfully completed or average time taken to complete a task). However, such participant observation is time consuming, costly and not easy to scale (Hargittai and Hsieh, 2012).

In this paper, we design and deploy an innovative measurement tool to create parsimonious and scalable survey measures for capturing digital literacy grounded in objective digital literacy measurements. Our measurement tool establishes a “gold standard” measure (or ground truth) of digital literacy through participant observation (PO). Each participant in our study completed a series of digital literacy tasks of increasing sophistication, which cover the essential steps needed to access and effectively find information online. It then uses a standard machine learning algorithm to capture the extent to which various survey measures predict the ground truth.[3]

Our study sample comprised 143 social media users in Pakistan, which included university staff (e.g., personnel involved with janitorial, classroom, and administrative services) and students of a university in urban Pakistan. Participants were recruited through a purposive sampling process, that is we sought to sample from populations with attributes that are likely to vary with digital literacy in the context of developing countries (e.g., low literacy, low-income, and females).[4] Thus, our sample included a significant fraction of low literate (44.6% of participants in our study received either no formal education or attained education below grade 6), low-income (at least 50% of participants in our sample had a monthly household expenditure below the national median household expenditure in Pakistan), and female participants (27.7% of our sample were females).

As PO studies are extremely challenging to scale due to the costs and labor involved, we evaluated the effectiveness of four (self-reported) survey modules for capturing the ground truth digital literacy of our sample. To this end, we assessed two *types* of survey questions: *platform-neutral* and *platform-specific* modules. We evaluated two platform-neutral modules, which included a module involving questions about the knowledge of Internet-related terms (which we refer to as the ‘Terms survey’), adapted from Hargittai and Hsieh 2012 and a survey module comprising basic digital literacy questions (‘Basic DL survey’) as well as two platform-specific modules, which included a module involving questions about the use of Facebook features (‘FB survey’) and another module comprising questions about WhatsApp features (‘WA survey’).

To find survey questions that are most predictive of the ground truth digital literacy, we use Random Forest (RF), a standard supervised machine learning method (Breiman, 2001). We use RF due to its effectiveness in capturing *non-linear* associations, flexibility, and robustness over small sample sizes (Biau and Scornet, 2016). This method has been employed in earlier studies on survey measures (e.g., Jayachandran et al. 2021; Guess and Munger 2020). Using RF, we obtain the most predictive survey modules by adding a constraint on the number of survey questions selected, a problem that is commonly referred to as *feature selection*.

We find that across the four survey modules, the 15-item Terms survey performed the best. This was followed by the Facebook features survey (‘FB survey’). We then constructed the best 7-item survey modules considering the following categories:[5] (i) all questions from the Terms survey (the best performing module out of the four), (ii) all platform-neutral questions, and (iii) all platform-specific questions. We find that the best 7-item platform-neutral module resulted in the lowest mean squared error (MSE) and highest R^2 values. This is followed by the best 7-item Terms survey, which is within 15.8% of the MSE and within 2.5% of R^2 of the best 7-item platform-neutral module. Even though platform-specific surveys ranked lower than the best 7-item platform-neutral module, they provide reasonable predictive power (e.g., the FB survey achieved a R^2 of 0.72). To further analyze their effectiveness, we implemented PO involving tasks related to specific Facebook and WhatsApp features. We find that performance on these tasks is predictive of performance on digital literacy tasks. Thus, such measures can be used by social media platforms, who already record users’ interactions with application features for providing functionality (e.g., searching on the platform).

We also analyzed how well demographic characteristics such as age, income, education, and employment status, predict the ground truth digital literacy. Unlike earlier studies involving U.S. or European samples (Sirlin et al. 2021, Guess and Munger 2022), we find income, education, gender, and employment status to be strong predictors of digital literacy. This likely points to barriers that different demographic groups face in improving their digital literacy due to lack of affordable Internet access, lack of affordable access to education, and lack of regular income (Qazi et al. 2021, Vashistha et al. 2019).

Our paper is part of the growing literature on measuring digital literacy which is important for bridging the digital divide (Dimaggio et al. 2004, van Deursen and van Dijk 2009, World Bank 2021). Closest to our work is Hargittai 2005 which used correlation between participant observation and self-reported familiarity of Internet terms for a sample of 100 internet users in New Jersey to propose survey-based digital literacy measures which were later updated in Hargittai (2009) and Hargittai and Hsieh (2012).[6] van Deursen and van Dijk 2009 conducted a PO study involving 109 Dutch citizens to complete nine government related assignments on the Internet. In another recent paper, Guess and Munger 2020 found that the survey measure by Hargittai (2005) to be most discriminating between respondents that were subjectively identified by the authors ex ante to possess “low-skills” ($N = 18$) or “high-skills” ($N = 83$).

Our work aims to bridge the knowledge gap in the measurement of digital divide, specifically with respect to digital literacy among low literacy populations and new technology users. To this end, we make several contributions to the existing literature. To the best of our knowledge, we are the first to conduct a PO study on digital literacy and carry out a validation of survey modules involving low skilled and hard to reach users in a developing country. Our research offers a scalable and cost-effective way for measuring digital literacy. This can be achieved by adding the recommended survey module to existing nationally representative household surveys carried out by governmental organizations annually. Social media platforms can also deploy

[3] Our approach draws upon the idea of *criterion validity*, which captures the extent to which a proposed measure agrees with a “gold standard” measure of the same construct (DeVon et al. 2007).

[4] Also known as the *fit-for-purpose* approach, it is widely used in political science (Guess and Munger, 2022) and pharmaceutical sciences (Lee et al. 2006) as a criterion guiding sampling; that is sampling from populations whose behavior is of considerable interest and likely to co-vary with the outcome of interest. Prior work shows that online convenience samples, such as those drawn from Amazon’s Mechanical Turk, may either miss or substantially undersample low digital literacy populations (Hargittai and Shaw, 2020; Guess and Munger, 2022).

[5] There is nothing special about *seven* questions, but this length seems appropriate for survey designers seeking a short instrument on digital literacy.

[6] The latter two works were surveys of Internet-related terms and did not involve a PO study. The study by Hargittai (2009) was conducted on a sample of students of an urban public university in the U.S. whereas the study (Hargittai and Hsieh, 2012) included two different samples of American users.

platform-specific survey items on their platforms to track the digital literacy of their users. Finally, developing scalable digital literacy measures with high validity can aid in tracking the distribution of digital skills across geographical regions, designing new interventions, user interfaces, and policies for improving digital literacy, and customizing interventions for users with different levels of digital literacy. For instance, such measures can be used for identifying populations who may be more vulnerable to negative externalities of technology, such as misinformation, and thus are important targets for interventions. This, in turn, can help reduce the spread and impact of misinformation and more generally, improve the informational well-being of people.

2. Participant recruitment and demographics

Our sample of 143 users comprised staff and undergraduate students of a university in Lahore, the second largest city by population in Pakistan. The university staff included personnel involved with janitorial services, classroom services, and administrative services. The survey and the participant observation (PO) were completed in two waves. In the first wave of the study, 43 subjects participated whereas in the second wave 100 subjects participated. While the first wave was conducted from August 23–25, 2021, the second wave was carried out from February 15–17, 2022. The participants were recruited via brochures that were placed throughout the campus as well as through purposive sampling. We offered an incentive of PKR 300 (USD 1.68) to each participant. Each survey (including the PO study) lasted for 40–45 mins. Our PO study was carried out over Android phones, which were provided by our team of surveyors. We used Android phones because of their popularity; over 3 billion active devices use Android worldwide (Cranz, 2022) as well as to ensure consistency in measurement. Further details about the study procedure are provided in Appendix Appendix A.3. The study was approved by the Institutional Review Board.

Rationale. The selection of our study sample was informed by the principle of '*fit-for-purpose*', which has been used in political science (Guess and Munger, 2022) and pharmaceutical sciences (Lee et al. 2006) as a criterion guiding sample selection, measurement, and validation efforts. For selecting study participants, this meant sampling from populations whose online behavior is of considerable interest and likely to co-vary with digital literacy in the context of a developing country (e.g., low-income, low literate, and female Internet users). Prior work shows that drawing (probability) samples from online platforms (e.g., Amazon's Mechanical Turk) may either miss or substantially undersample low-skill populations (Hargittai and Shaw, 2020; Guess and Munger, 2022). For measurement, the fit-for-purpose approach meant using specific tasks, such as the ability to find information online, as competencies to be predicted or explained. Finally, for validation, this meant drawing upon the idea of *criterion validity*, which captures the extent to which a proposed measure agrees with a "gold standard" measure of the same construct (DeVon et al. 2007).

Participant demographics. 72.3% of participants in our study were males whereas 21.7% were females. 65.5% of the participants were below the age of 29 (median age was 26 years) and 25% were between 33–55, and the maximum age was 55 years. The age distribution in our sample is comparable to the national distribution in Pakistan where 64% of citizens are estimated to be below the age of 29 and nearly 28% were between 30–55 (compared to 34.5% in our sample). The national median age is 22.5 years (Najam and Bari, 2018).[7] 21.7% of our sample received no formal education, 23.1% had received formal education between grade 1 and grade 6, 28.7% between grade 6 and 12, and 26.5% above grade 12.[8] 24.7% of the participants had a monthly household expenditure under PKR 30,000 (USD 167.8), 46.1% between PKR 30,000 and PKR 70,000 (USD 391.2) and 29.2% above PKR 70,000. According to the Household Integrated Economic Survey 2018–19 conducted by the Pakistan Bureau of Statistics, the median monthly household expenditure in urban areas in Pakistan was estimated to be PKR 31,031. After adjusting for inflation, we estimate this amount to be PKR 41,987 in 2022 Pakistani Rupees, which is similar to the median household expenditure (i.e., PKR 40,000) in our sample. 79% of the sample had a full-time employment status, 11.2% were employed part-time, and 9.8% were students (and were not employed). Appendix Table A.1 provides details about the descriptive statistics of our sample.

[7] However, one difference is that while 8% of the national population is estimated to be older than 55 years, our sample does not include individuals from this category.

Social media use. All participants in our study were social media users and all of them had a WhatsApp account. 96.5% reported using their own mobile phones for accessing social media whereas the rest did not use their own device. The percentage of users in our sample having Facebook, YouTube, Instagram, and Twitter accounts was 87.4%, 93.7%, 46.2%, and 29.4, respectively. This distribution of accounts across social media platforms is similar to the finding of an earlier study based on a random sample from the city of Lahore, which found that WhatsApp was the most popular platform followed by Facebook and Twitter (Ali and Qazi, 2021). Facebook was the primary source of news for the largest fraction of participants in our sample (30%) followed by TV (21.7%). Interestingly, for the remaining 48.3% of users – who did not use Facebook or TV as their primary source of news – reported using Facebook (25.2%), TV (16.8%), and WhatsApp (15.4%) as the second primary source of news. This highlights the importance of social media platforms as an important source of receiving news.

3. Measuring ground truth of digital literacy

To measure the ground truth of digital literacy (DL), we draw upon the tradition of conceptualizing DL as the ability to effectively *find* information online (Hargittai, 2005; Gilster, 1997).[9] Viewed this way, DL comprises two key components: (i) an *information literacy* component, which refers to the literacy required to find relevant information online effectively (e.g., looking up answers to questions and verifying claims using various strategies) and (ii) a *digital skills* component, which comprises a set of basic digital knowledge and competencies needed to use the Internet and digital media and attain information literacy. For example, finding information related to a news headline encountered on social media would require the ability to read and understand the news content, extract relevant information, open a browser, access a search engine (e.g., by typing its URL), and enter a suitable query. Thus, the ground truth DL can be measured by asking people to complete a series of tasks that capture the steps needed to find information online and observing their effectiveness in completing these tasks (Hargittai, 2005; van Deursen and van Dijk, 2009).

We establish the ground truth DL using in-person observations.[10] Each participant in our study completed a series of DL tasks of increasing sophistication, which resulted in two objective measures of online

[8] 29% of Pakistan's population is non-literate (i.e., they received no formal education) whereas 65% received education between grade 1 and grade 12 (compared to 51.8% in our sample).

[9] Recently, Guess and Munger 2020 postulated digital literacy to mean online *information discernment*, i.e., the ability to reliably assess the credibility of information encountered online. While the conceptualization we draw upon is different as it does necessarily *require* discernment, we include multiple tasks in our participant observation study that is related to information discernment.

[10] Another possible approach for capturing the ground truth DL is to use automated, remote tests. For example, an online "quiz" could directly assess some of the skills mentioned in Table 1 (e.g., participants could be asked to search for a term on Google and copy and paste the result). However, there are limitations of what a browser can reliably detect with respect to user's online behaviors (e.g., ability to connect to a WiFi network, making bookmarks, opening new tabs, clearing cache, etc.). Moreover, obtaining such information directly from the browsers may raise privacy concerns.

Table 1
Digital literacy tasks for ground truth measurement using in-person observations and their completion rate (i.e., proportion of participants who successfully completed the task).

Digital literacy task	Completion rate (N = 143)
1. Connect to a WiFi network	0.87
2. Search the term "LUMS" on google.com	0.82
3. Open a mobile browser	0.78
4. Look up the birthplace of "Quaid-e-Azam"	0.73
5. Open a new tab in the browser	0.52
6. Observe a news headline on social media and find information relevant to the headline using a search engine. (*participants were shown a screenshot of the news headline and asked to find information relevant to the content in the screenshot*)	0.51
7. Copy the URL that appears after searching "LUMS"	0.45
8. Bookmark a webpage	0.34
9. Clear all cache and cookies from your browser	0.30

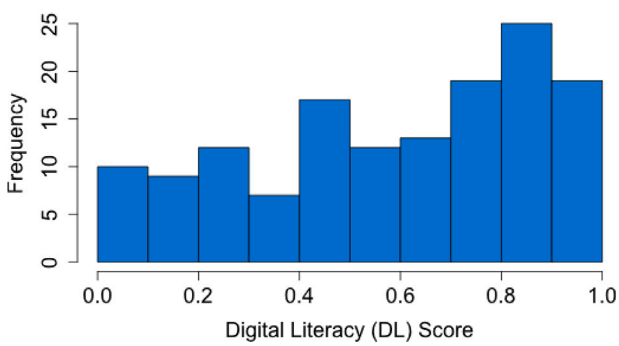

Fig. 1. Frequency plot of ground truth digital literacy scores of 143 study participants. The DL score of a participant is defined as the proportion of successfully completed tasks out of a total of nine DL tasks. The mean and median DL scores in our sample were 0.59 and 0.67, respectively.

skill: (i) the percentage of tasks successfully completed (*effectiveness*) and (ii) the amount of time spent on the tasks (*efficiency*). The set of DL tasks we use in our study cover the essential steps needed to effectively find information online and thus included tasks that capture users' computer/mobile and web-use skills (e.g., the ability to connect to WiFi and open a browser) as well as their information literacy (e.g., ability to search for the birthplace of a personality, ability to observe a news headline and being able to find relevant information online) as shown in Table 1.

To capture the high-end of DL (i.e., users who are both highly effective as well as efficient in finding information), we include the ability to bookmark a webpage as a task in our set as it is related to being able to organize information and gain fast access to it. Similarly, the ability to clear cache and cookies from a browser is related to understanding of user privacy (e.g., how one's information may be used to show personalized content or ads) and how can it be managed. Finally, the ability to open a new tab on a mobile browser can capture the skills needed to multi-task and access multiple sources of information quickly.

Some tasks were intentionally designed to include redundancy in them (e.g., *Search the term "LUMS" on www.google.com* and *Look up the birthplace of "Quaid-e-Azam*[11]") because they allow us to capture more granular difference in DL. For instance, in the former task, a user needs to enter a given term "LUMS" on www.google.com whereas in the latter task, a user needs to compose a suitable query from the provided information and then find the birthplace from the search results. The difficulty of the latter task is also reflected in its lower completion rate (e.g., 0.73 vs. 0.82) as shown in Table 1.

In general, we observe that the completion rate of tasks decreases with the increase in task sophistication; see Table 1. Thus, the ability to clear browser cache and cookies had the lowest completion rate whereas the ability to connect to a WiFi network had the highest completion rate.[12]

Using data on task completions from our PO study, we calculate the ground truth DL score of a participant to be the completion rate across all the nine tasks (i.e., the proportion of tasks successfully completed by the individual).[13] The mean and median DL scores were 0.59 and 0.67, respectively. Fig. 1 shows the frequency distribution of DL scores in our study sample of 143 users. We found that the respondents varied considerably in their capacity to access, understand, and find information online. For example, 25% of participants in our study could only complete up to 3 tasks out of a total of 9 tasks whereas the top 30% completed between 7 to 9 DL tasks.

4. Survey measures of digital literacy

It is extremely challenging to scale PO studies due to the cost and labor associated with conducting them, which makes it important to design self-reported survey modules that can serve as reliable proxies for measuring the ground truth DL (Hargittai 2005).

An ideal survey module to serve as a proxy for the underlying DL will have several properties. First, it should be correlated with the ground truth (Hargittai and Hsieh, 2012). Second, it should capture variations at both the low-end and high-end of the DL spectrum. Third, it should reflect the *process* (e.g., knowledge of digital spaces, digital skills) that causes the effects of Internet and digital media use to vary across individuals (Guess and Munger, 2022).

To find survey-based proxies that are most predictive of the ground truth, we evaluated the effectiveness of two types of survey modules: (i) *platform-neutral* modules, which comprise questions that do not test knowledge or use of any specific social media platform[14] and (ii) *platform-specific* modules, which include questions that test the knowledge and/or use of features of social media platforms (e.g., Facebook and WhatsApp).

[11] Quaid-e-Azam (or the *great leader*) refers to the founder of Pakistan.

[12] Observe that the completion rate of task 2 is higher than task 3, which may seem counter-intuitive because presumably the latter task should be a required to complete task 2. However, we noticed that a small fraction of users who could not open a mobile browser but were able to search the term "LUMS" on google used the Google voice assistant available on Android phones.

[13] Our efficiency analysis of PO tasks (i.e., the time taken to complete a task) shows that users with higher digital literacy are able to complete tasks more quickly than lower digital literacy. As expected, we observe that the variance in task completion times grows smaller for harder tasks, due to their lower completion rates; see Appendix A.2.

[14] In general, "platform" may refer to any digital platforms (e.g., operating system, databases) but we use 'platform' to refer to social media platforms only.

Table 2
Mean Squared Error and R^2 for different survey modules on the training and Out-of-Bag (OOB) samples using a Random Forest regression model.

Survey Modules	Items	MSE (OOB)	R^2 (Training)	R^2 (OOB)
1. Basic Digital Literacy Module	7	0.048	0.60	0.52
2. Internet-related Terms Module	15	**0.022**	**0.94**	**0.78**
3. Knowledge of Facebook Features Module	9	0.039	0.72	0.56
4. Knowledge of WhatsApp Features Module	9	0.056	0.55	0.45
Best 7-item Survey Modules				
α. Best 7-item Internet-related Terms Module	7 (/15)	0.022	0.89	0.78
β. Best 7-item Platform-neutral Module	7 (/22)	**0.019**	**0.90**	**0.80**
γ. Best 7-item Platform-specific Module	7 (/18)	0.033	0.73	0.63

Platform-neutral modules are attractive because they can be used to compare the DL of users across populations (e.g., users of different social media platforms, individuals who do not use social media). Platform-specific survey measures are useful for three key reasons. First, social media applications – such as Facebook and WhatsApp – have become extremely popular especially in developing countries and make up a significant fraction of digital media consumption of Internet users (DataReportal, 2022). Second, it is common for social media platforms to provide features for: (i) finding different types of content (e.g., finding pages on Facebook, searching for chat messages on WhatsApp), (ii) understanding the type of content (sponsored vs. non-sponsored content on Facebook), and (iii) managing one's personal information (e.g., via use of privacy settings). Thus, the use of such platform features reflects the ability to effectively find information and therefore captures the process by which DL may vary across individuals. As a result, the ability to use platform features is likely to be correlated with the ground truth DL. Finally, a desirable aspect of such platform-specific questions is that their responses can be implicitly gleaned from users' interactions with the platform without requiring them to fill a survey.

We evaluate and analyze four survey modules, which include two platform-neutral and two platform-specific modules:

1. *Internet-related Terms Module ('Terms Module')*: This platform-neutral module measures self-reported familiarity with Internet-related terms (e.g., browser, bookmark).
2. *Basic Digital Literacy Module ('Basic DL Module')*: This platform-neutral module measures self-reported ability to accomplish various DL tasks.
3. *Knowledge of Facebook Features Module ('FB Module')*: This platform-specific module measures self-reported knowledge and use of various Facebook application features.
4. *Knowledge of WhatsApp Features Module ('WA Module')*: This platform-specific module measures the self-reported knowledge and use of various WhatsApp application features.

4.1. Comparison of survey modules

To compare the effectiveness of the four survey modules in predicting the ground truth DL, we train four regression models – one for each survey module – using Random Forest (RF); a standard machine learning algorithm that can efficiently fit non-linear relationships in the data, provides robustness performance over small sample sizes (Breiman, 2001; Biau and Scornet, 2016) and can be applied when predictor variables are highly correlated (Strobl and Zeileis, 2008). In each model, the outcome variable is the ground truth DL score whereas the independent variables are the responses to survey questions. We ran each RF model with 100,000 trees and used the default values for the tuning parameters.

Table 2 shows the results for the four surveys modules in terms of three metrics: Mean Squared Error (MSE) over Out-of-Bag (OOB) samples,[15] R^2 over the training samples, and R^2 over the OOB samples. Observe that across all the three metrics, the Terms module achieves the best performance (i.e., the smallest MSE and the largest R^2 values) indicating its effectiveness in predicting DL. This is followed by the FB module, Basic DL module, and the WA module each of which exhibit moderate predictive performance. The R^2 values for these survey modules suggests that they can also serve as useful proxy measures for the ground truth DL albeit with less accuracy.[16]

Next, we trained models for shorter 7-item modules drawn out of the Terms module, platform-neutral module (which combines questions from the Terms and Basic DL modules), and the platform-specific module (which combines questions from the FB and WA modules).[17] We find that the 7-item platform-neutral module performs the best across all modules. The 7-item Terms module was the next best, which was within 16% of the best module in terms of MSE and within 2.5% in terms of R^2 computed over OOB data.

4.2. Platform-neutral modules

We now analyze each platform-neutral survey module separately. In particular, we evaluate the relative importance of each question in a module in terms of their contribution to the predictive power of the survey module.

4.2.1. Knowledge of internet-related terms module

We now examine the 15-item Terms survey, which was first proposed by Hargittai 2005. In this survey, respondents are asked to self-report their familiarity with computer- and Internet-related terms using the following question: *How familiar are you with the following computer and Internet-related items? Please choose a number between 1 and 5, where 1 represents no understanding and 5 represents full understanding of the item.*

The 15-item survey we use included Internet-related terms previously examined (Hargittai 2009; Hargittai and Hsieh 2012) as well as four new terms (i.e., "Internet", "MB/GB", "MP4", and "URL") as shown in Table 3. These items are designed to reflect terms that are common in everyday Internet use, such as "Browser", "Search engine", and "URL", as well as more technical terms but presumably less common, such as "Cookies" and "Malware". This approach makes sense as a way to capture a spectrum of digital skills and is also flexible because the items can be adapted to include more recent terms.

We also included two bogus terms in our survey to see if participants were responding randomly or were affected by social desirability bias. Thus, if the bogus terms score significantly better than or as well as the real terms then we must question the validity of the measure as a proxy

[15] OOB samples are bootstrap samples that are not used for training the model. Instead, they are used for evaluating how well the model generalizes (James et al. 2013).

[16] In Appendix Table A.3, we also report the Pearson's correlation coefficient and Spearman's correlation coefficient for the survey modules.

[17] See Online Appendix Table 6 for an extended set of results including those involving other combinations of survey modules to serve as proxies for predicting digital literacy.

Table 3
Column 2 shows the average score (or rating) for each Internet-related item across respondents. Columns 3 and 4 show two metrics of variable importances with random forest: %IncMSE and IncNodePurity, respectively. The terms with the ⋄ symbol are part of the best 7-item Terms module whereas the terms with the • symbol are part of the best 7-item platform-neutral module. $^{*}p < .10$;$^{**}p < .05$;$^{***}p < .01$. $N = 143$, R^2 (OOB samples) = 0.78.

Term	Average score	%IncMSE	IncNodePurity
1. Internet ⋄•	3.80	261.4**	1.33**
2. MP4	3.69	125.1	0.47
3. Browser ⋄•	3.55	261.4***	0.96**
4. Search engine	3.46	160.9*	0.89
5. MB/GB	3.40	156.2	0.34
6. PDF ⋄•	3.10	320.1***	3.68***
7. Bookmark ⋄•	2.93	213.6**	1.85***
8. JPG ⋄	2.69	137.2**	1.32***
9. URL ⋄•	2.66	151.8*	0.94**
10. Cookies	2.52	175.8**	0.33
11. Torrent ⋄•	2.45	142.1**	0.63**
12. Podcasting	2.16	146.9	0.29
13. Malware	2.01	115.1	0.24
14. Phix (bogus)	1.45	–21.5	0.07
15. Jcrypt (bogus)	1.29	–26.4	0.07

for DL because it suggests that respondents are not reporting levels of understanding if they claim to understand bogus terms.

Table 3 shows the mean scores of the terms. First, the two bogus items had the lowest mean score, therefore, it is reasonable to assume that people did not select their responses randomly and that the module measures people's understanding of Internet-related terms. Second, the wide range of mean scores across the terms indicates that they are able to capture a spectrum of ground truth DL.

Next, we trained a RF regression model with each term added as a distinct feature to analyze their contribution to the predictive power of the Terms survey module. To evaluate feature importance, we employ two commonly used metrics for evaluations involving RF regression: *%IncMSE* and *IncNodePurity*. The former metric is the percentage increase in MSE when that feature variable is removed from analysis[18] and the latter metric is the total decrease in node impurities (or residual sum of squares) from splitting on a feature variable in the decision tree, averaged over all trees used in the RF (Breiman, 2001; Biau and Scornet, 2016).

Table 3 also shows the importance of each term in predicting ground truth DL along with their p-values.[19] Observe that the terms PDF, Internet, Browser are the top three terms in terms of the magnitude across both the metrics. Observe that values for these metrics are statistically significant at the 5 percent level or less (i.e., $p < 0.05$). Again, the bogus terms have the smallest values across both the metrics, which indicates that respondents did not know about these terms and thus were highly unlikely to have responded randomly. In Section 7, we show that *different* terms best predict DL of users at the low and high end of the digital literacy spectrum.

4.2.2. Basic digital literacy module

Next, we evaluate a basic DL survey comprising questions shown in Table 4. These questions aim to capture the skills and literacy needed to access, understand, and find information online through self-reported responses.

We find that the questions, "*Are you able to search/google things online?*", "*Are you able to read text on social media?*", and "*Are you able to connect to WiFi and/or turn on mobile data?*" as the best three predictors in this survey module across both metrics (see Table 4). The value of both metrics for these questions are statistically significant at the 1 percent level. Indeed, the ability to find information on a search engine is an important information retrieval task whereas being able to connect to WiFi and/or mobile data and read text capture users' ability to access the Internet and read and understand content online. The latter two questions are particularly helpful in separating users at the lower end of DL (see Section 7).

4.3. Platform-specific modules

Platform-specific surveys can be useful measures for the ground truth DL for a number of reasons. First, social media applications provide features whose use can capture a range of digital skills as well as information literacy (e.g., searching on the platform, updating privacy settings of your account and reporting a post on Facebook). Second, social media platforms provide features (or clues) to potentially spot low quality content (e.g., discerning between sponsored and non-sponsored posts). Knowledge of these features can be suggestive of information discernment. We now analyze the importance of individual items in the FB and WA survey modules.

4.3.1. Facebook module

In the FB survey, we asked participants about their use of nine Facebook application features; see Table 5. For each question, we construct a dummy variable (equal to 1 if a participant uses the feature and 0 otherwise).[20] We find that the top three predictors were the ability to update the privacy settings of one's account and posts, and the ability to identify which posts are sponsored or not (see Table 5).

4.3.2. WhatsApp module

Next, we analyze the WA survey module, which also comprised nine items. Again, for each feature we construct a dummy variable (equal to 1 if the they use the feature and 0 otherwise). Our RF regression results show that the top three predictors include: (a) the ability to report a user, (b) ability to determine whether a sent message has been seen, and (c) the ability to block a user all of which are statistically significant at the 1 percent level (see Table 6).

Overall, the WA survey had the highest MSE and lowest R^2 across the four modules, which suggests that it may not be as effective as other modules.

In summary, the FB and WA modules can serve as useful proxies for capturing the underlying DL. However, such measures are generally only useful for measuring the DL of users of a given platform as they may not allow for cross-platform comparisons due to differences in the use of social media platforms across regions. For example, in our study sample WhatsApp was more popular than Facebook and Twitter was the least popular platform in terms of usage. More generally, the effectiveness of platform-specific surveys for measuring DL depends on how well the application features test digital skills and information literacy.

[18] This metric – which is also known as *permutation importance* – breaks the relationship between the feature and the outcome variable by randomly shuffling feature values using OOB data (Altmann et al. 2010). As a result, the increase in MSE is indicative of how much the model depends on the feature.

[19] We compute the p-values for feature importance metrics by permuting the response variable, which produces a null distribution for each predictor variable. For this purpose, we use the rfPermute package in R (Archer, 2016). For each survey module, we train a RF model using 100,000 trees and conduct 10,000 repetitions for finding the p-values. We use the default values for the rest of the parameters.

[20] 87.4% of the respondents (125 out of 143) in our study were Facebook users. The mean DL score in this sample was 0.625, which is about 6% higher than the mean DL score of the entire study sample.

Table 4
Column 2 shows the average score (or rating) for each basic digital literacy question across respondents. Columns 3 and 4 show two metrics of variable importances with random forest: %IncMSE and IncNodePurity, respectively. The item(s) with the • symbol are part of the best 7-item platform-neutral module. $^{*}p < .10$;$^{**}p < .05$;$^{***}p < .01$. $N = 143$, R^2 (OOB samples) = 0.52.

Survey module question	Options	Avg. score	%IncMSE	IncNodePurity
1. Are you able to connect to WiFi or turn on mobile data?	*Yes; No*	0.85	271.1***	1.89***
2. Are you able to read text on social media?	*Yes; No*	0.83	392.6***	1.92***
3. Are you able to use social media applications generally without assistance?	*Yes; No*	0.92	209.1**	0.49*
4. Are you able to search/google things online? •	*Yes; No*	0.81	418.6***	2.89***
Please select all the activities you perform on social media:				
5. Viewing posts; messages; images; videos; news articles	✓; ☐	0.99	0.0	0.02
6. Creating a post; message; image; video	✓; ☐	0.80	44.1	0.25
7. Sharing a post; message; image; video; news article	✓; ☐	0.80	−24.7	0.19

Table 5
Column 2 shows the average score (or rating) for each Facebook feature across respondents. Columns 3 and 4 show two metrics of variable importances with random forest: %IncMSE and IncNodePurity, respectively. The item(s) with the • symbol are part of the best 7-item platform-specific module. $^{*}p < .10$;$^{**}p < .05$;$^{***}p < .01$. $N = 125$, R^2 (OOB samples) = 0.56.

Facebook feature	Avg. score	%IncMSE	IncNodePurity
1. Create post	0.68	35.0	0.28
2. Like a post	0.87	0.0	0.14***
3. Comment •	0.78	166.2*	0.45*
4. Update privacy settings of your account •	0.58	322.3***	2.1***
5. Update privacy settings of posts •	0.55	260.5**	2.0***
6. Report a post	0.48	18.0	0.25
7. Hide an ad •	0.44	232.9**	1.11***
8. Hide a post	0.45	93.2	0.54*
9. Identify sponsored vs non-sponsored posts •	0.48	267.3**	1.2***

Table 6
Column 2 shows the average score (or rating) for each WhatsApp feature across 143 respondents. Columns 3 and 4 show two metrics of variable importances with random forest: %IncMSE and IncNodePurity, respectively. The terms with the • symbol are part of the best 7-item platform-specific module. $^{*}p < .10$;$^{**}p < .05$;$^{***}p < .01$. R^2. $N = 143$, R^2 (OOB samples) = 0.45.

WhatsApp feature	Avg. score	%IncMSE	IncNodePurity
1. View chats	0.97	−0.92	0.14
2. Reply to a chat	0.89	170.0	0.59***
3. Record audio messages	0.97	6.3	0.16
4. Forward message	0.91	73.6	0.45***
5. Delete message	0.94	−38.2	0.12
6. Report user •	0.55	414.6***	2.55***
7. Block user	0.80	267.0***	1.56***
8. Message seen •	0.88	354.4***	1.69***
9. Audio note speed change	0.71	56.2	0.37

4.3.3. Participation observation study of platform-specific DL tasks

On a sub-sample of participants in our study, we conducted a PO study to measure the actual ability to use various Facebook and WhatsApp features. This is particular attractive because social media platforms often keep track of user interactions with application features. Thus, if such measures predict ground truth DL, they can be conveniently employed as a proxy measure by platform providers. In turn, this can inform digital literacy based interventions to counter misinformation (Sirlin et al. 2021; Ali and Qazi 2021).

The completion ratios for Facebook and WhatsApp tasks are shown in Table 7. Observe that they cover a wide range of completion rates. Moreover, these PO tasks are able to more precisely measure the ability to use features than through self-reported responses via the FB survey module even though the difference is not large.

FB tasks. We find that the five Facebook tasks predict the ground truth DL well with a MSE and R^2 values (over OOB samples) of 0.037 and 0.66, respectively. Interestingly, our subsample had a mean ground truth DL score of 0.53, which is nearly 10% lower than the mean DL score of the entire sample, therefore, the subsample is skewed towards low-end DL users. Taking a closer look at the individual questions, we find that Facebook tasks 3–5 (Table 7) were most predictive of DL scores as shown in Fig. 2. These include searching for the 'LUMS' facebook page, following the 'LUMS' facebook page and changing the privacy settings of a post. These tasks relate to being able to find information online.

WA tasks. In case of WhatsApp tasks, the MSE was higher (0.055) and R^2 was lower (0.53) compared to Facebook tasks, which suggests that they are less predictive of DL scores even though the mean DL score in this subsample was also 0.53. The most predictive tasks include the ability to mute notifications for a chat, block a contact, and send a chat message as shown in Fig. 2.

5. Best 7-item modules and indices

Employing survey modules with large number of questions just for measuring DL poses a challenge for studies where the primary focus is not on understanding DL and thus have less space for such questions. To address this concern, we find shorter 7-item survey modules and evaluate their effectiveness for serving as proxies for the ground truth DL.

To find the best 7-item survey modules, we consider the following categories of questions: (i) a 15-item Term module (the best performing module out of the four survey modules), (ii) a 22-item module that combines questions from the two platform-neutral surveys ('Platform-neutral module'), and (iii) a 18-item module that combines questions from the FB survey and WA survey ('Platform-specific module'). Note that we do not include platform-specific *participant observation tasks* as they reduce the sample size substantially.

To find the best 7-item modules, we select the top seven items from each category of larger modules using the following criterion: We first select all items that are statistically significant at 10 percent or less (i.e., having a p-value less than 0.10) across *both* the metrics (i.e., %IncMSE and IncNodePurity), we then sort them by the magnitude of the %IncMSE metric, and finally select the top seven items. We select based on the magnitude of %IncMSE because it is considered a more robust measure of variable importance (Strobl et al. 2007).

We find that the resulting best 7-item platform-neutral module included the following items: *PDF, Bookmark, Internet, Browser, URL, Torrent, 'Are you able to search/google things online?'*. While six items were from the Terms module, one item was from the Basic DL module.

Table 7
Results of a participant observation study involving Facebook tasks (N = 34) and WhatsApp tasks (N = 44).

Facebook task	Completion rate
1. Like/React to a post	0.97
2. Comment on a post	0.84
3. Search for LUMS facebook page	0.78
4. Follow LUMS facebook page	0.68
5. Change the privacy settings of a post on your timeline to 'Only Me'	0.35
WhatsApp task	
1. Send a chat message (on a given number)	0.90
2. Send a voice note	0.92
3. Block a contact on WhatsApp	0.79
4. Mute notifications for a chat	0.46

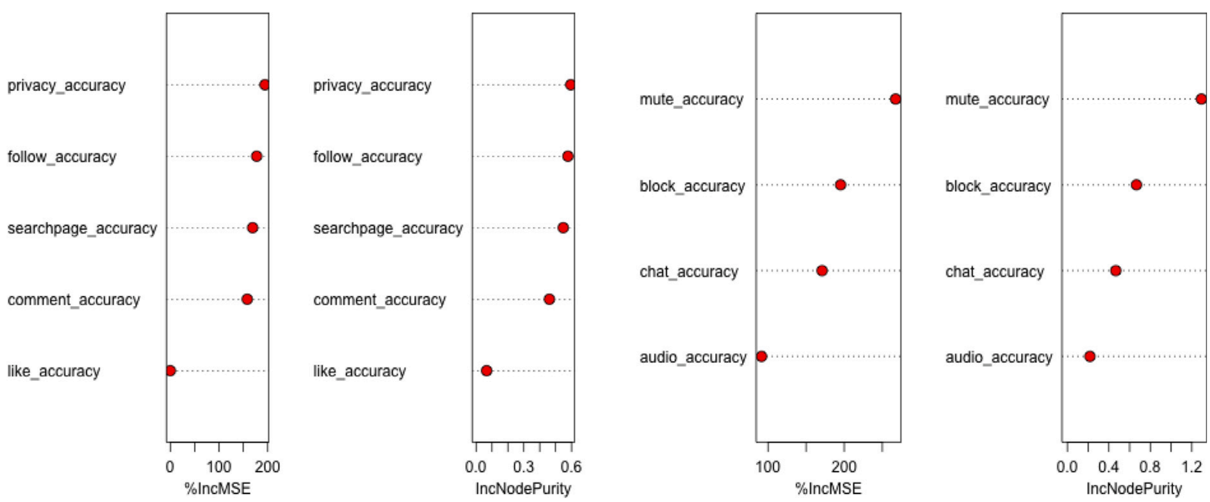

Fig. 2. Results of a Random Forest regression. (i) The two leftmost plots correspond to Facebook tasks (N = 34) and (ii) the two rightmost plots correspond to WhatsApp tasks (N = 44).

This was the best performing module across the categories we considered and resulted in a MSE of 0.019 and a R^2 value of 0.80 over OOB data.

Next, we find the best 7-item platform-specific module. It included five items from the FB module (items 3, 4, 5, 7, and 9 shown in Table 5) and 2 items from the WA module (items 6 and 8 in Table 6). The resulting MSE and R^2 were better than any of the individual surveys.[21] Such a module can be launched for measuring the DL of users of multiple platforms (e.g., Facebook and WhatsApp). For entities like Meta, which owns Facebook, WhatsApp, and Instagram, this may be particularly feasible.

We find that the best 7-item Terms survey module included the following terms: *PDF, Bookmark, Internet, Browser, JPG, URL, Torrent* with a MSE = 0.022 and R^2 equal to 0.78 over OOB data, which is as almost good as using the 15-item Terms module. Moreover, 7-item Terms module is only 15.7% higher in MSE and 2.5% lower in R^2 (OOB data) compared to the best performing module (i.e., the platform-neutral module).[22]

6. Demographic correlates of digital literacy

We collected data about the demographic characteristics of our study sample, which included data on age, gender, education, income, and employment status of participants.[23] We now analyze (i) the correlation between these demographic characteristics and the ground truth DL and (ii) their effectiveness in predicting the DL relative to the four individual survey modules we evaluated.

Fig. 3 shows the coefficients of a OLS regression model, where the dependent variable is the ground truth DL scores and the independent variables are the demographic characteristics. We find that education, household income (for which we use household expenditure as a proxy), employment,[24] and gender have large and statistically significant (at the 10 percent level) effect sizes. The predictive power of these characteristics highlights the importance of taking into account cultural context in studies on digital literacy. It also points to potential barriers people face in improving their digital literacy due to lack of affordable

[21] In Appendix A.4, we report the distribution of DL scores predicted by the Best 7-item platform-specific survey and the Basic DL survey on a sample (N = 618) of Facebook users in Lahore, Pakistan, whose data was collected in an earlier study in 2019. We find that both survey modules are able to detect significant variations in digital literacy.

[22] In Online Appendix E, we report the best 7-item survey modules constructed from the Terms survey, platform-neutral survey, platform-specific survey, and global survey (which included items from all of the individual survey modules we evaluated) as well as feature importances of each survey item along with their corresponding p-values.

[23] See Online Appendix Table 3 for details about response options and coding of responses.

[24] The negative association between employment and DL scores can be explained by the fact that the only unemployed individuals in our sample were students, who had higher DL scores than employed individuals on average.

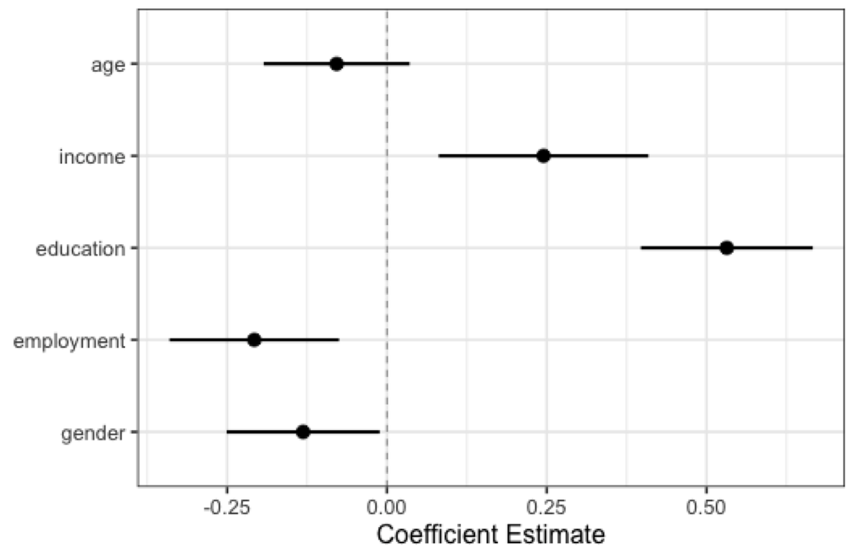

Fig. 3. Results of OLS regression. The outcome variable is the DL score of respondents and the independent variables are the five demographic characteristics of participants. All variables are standardized. The plot shows regression coefficients along with 90% confidence intervals.

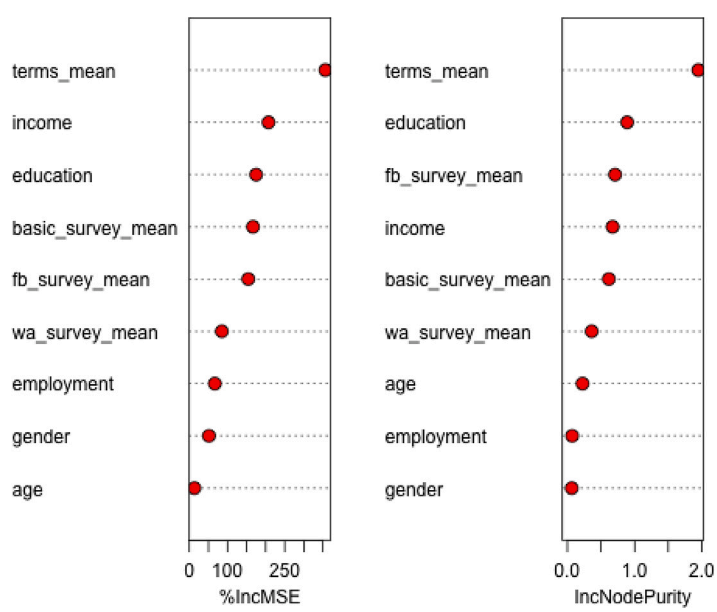

Fig. 4. Results of a Random Forest regression. Each row in the left panel shows the percentage increase in MSE when that feature is removed from analysis, while each row in the right panels shows the average increase in node purity. $N = 89$.

access to the Internet, lack of affordable access to education, and lack of regular income (Qazi et al. 2021, Vashistha et al. 2019). The negative correlation between gender scores and DL scores is also suggestive of the digital gender divide that exists in several developing countries. For example, currently Pakistan ranks the highest in the world in terms of the gender gap in Internet access between men and women (65%) and mobile phone ownership (51%) (EIU Inclusive Internet Index, 2021). It is useful to note that data on some of these characteristics may be challenging to obtain due to privacy considerations. For example, in our study, only 89 participants responded to the question about monthly household expenditure.

To compare the relative effectiveness of demographic characteristics and other survey modules in predicting DL of respondents, we conducted a 'horserace' using a RF regression model in which we added the mean scores by each of the four survey modules and each demographic characteristic as a separate feature (see Fig. 4). While the results suggest that adding demographic variables may improve the accuracy of predicting DL in a study sample similar to ours, however, this insight is unlikely to generalize to other populations. The reason is that demographic characteristics do not directly capture the knowledge about digital spaces, digital skills or information literacy. Rather, they serve as a proxy for some other characteristics that relate to people's experiences online. Thus, they likely point to barriers that different demographic groups face in improving their digital literacy (e.g., less educated, low-income, females, and individuals with part-time employment in our study sample). For example, studies conducted over U.S. and European Internet users have found age to be correlated with DL (Sirlin et al. 2021, Guess and Munger 2020). However, in our study sample where lower income levels and affordable access to Internet and education are significant barriers to improving one's DL, we do not find age to be a strong predictor of DL (see Fig. 3).

Table 8
Mean Squared Error and R^2 (Training Data, OOB data) for Low DL (L) and High DL (H) users across two survey modules using a Random Forest regression model. The mean DL score in the low DL and high DL groups were 0.30 and 0.85, respectively.

Survey modules	MSE (L)	MSE (H)	R^2 (L)	R^2 (H)
1. Best 7-item Platform-neutral Module	0.016	0.010	(0.78, 0.56)	(0.55, 0.21)
2. Basic DL Module	0.027	0.014	(0.51, 0.26)	(0.067, −0.046)

7. Separating low DL and high DL users

Our study sample comprised participants with a wide range of digital literacy. To understand why the best 7-item platform-neutral survey module performs well, we divide our sample into a low DL group and a high DL group and analyze its effectiveness in predicting the ground truth DL within each group. To construct the two DL groups, we divide our sample by the median DL score. For comparison purposes, we consider the Basic DL module. We train RF regression models for the best 7-item platform-neutral module and the Basic DL module. For each module, we trained one RF model for the low DL group and one for the high DL group. We make the following observations:

- We found that while the platform-neutral module was effective at separating individuals within each sub-group group, the Basic DL module was only effective at differentiating between users in the low DL group (see Table 8).
- For low DL users, the term ‘*Internet*’ and ‘*Are you able to search/google things online?*’ were the top two predictors of DL whereas for high DL users, the terms *URL* and *Torrent* were most predictive (in terms of %IncMSE). This indicates that the platform-neutral module is able to capture variations at both the low-end and high-end of the DL spectrum by having terms/questions with a range of difficulty.
- With the Basic DL survey, the ability to search online (item 4 in Table 4) and the ability to connect to WiFi or turn on mobile data (item 1 in Table 4) were the best two predictors in the low DL group. However, none of the questions are effective at measuring digital literacy of users in the high DL group as it does not include more difficult questions.

In summary, the predictive power of the platform-neutral module can be attributed to its effectiveness in capturing variations at both the low-end and high-end of the DL spectrum by having questions of varying difficulty (or sophistication).

8. Recommendations

We recommend using the best 7-item platform-neutral survey module, which resulted in the lowest MSE and the highest R^2 values across all categories of survey modules we considered in this work (see Table 9). For instance, it resulted in the highest R^2 (0.8) over the OOB data, which points to its effectiveness in generalizing to unseen samples.[25] To map survey responses to digital literacy scores, we have made our trained RF model publicly available at the following link: https://github.com/nsgLUMS/predict_DigitalLiteracy. The model takes as input one or more observations, where an observation comprises responses to the 7-item platform-neutral survey module, and predicts the individuals' digital literacy score.

Fig. 5. Accuracy achieved by the best 7-item platform-neutral module in predicting quintile ranks of the ground truth DL score. The plot shows the confusion matrix, where the diagonal shows the probability of correctly predicting a quintile rank (e.g., rank 2 was correctly predicted with probability 0.91).

To further aid researchers and practitioners in using the recommended survey module, we assess its accuracy in predicting the quintile ranks of the ground truth DL scores.[26] Fig. 5 shows the confusion matrix for the best 7-item platform-neutral survey. Observe that the module achieves accuracy between $[0.78, 0.96]$ for the first four quintile ranks (i.e., ranks corresponding to the bottom 80th percentile of DL scores) but the accuracy reduces to 0.5 for the fifth quintile, which comprises users with the highest 20% DL scores in our sample.[27] To interpret the accuracy of the fifth quintile rank, two observations can be instructive: (i) a random classifier will achieve an accuracy of 0.2 in correctly predicting the last rank whereas the platform-neutral survey improves on such a classifier by 2.5× and (ii) the predicted rank is either 4 or 5 with an accuracy of 0.82, thus most mis-classifications are clustered around 5.

9. Conclusion

Digital literacy can play an important role in expanding economic opportunities by enabling people to effectively find and consume valuable information online, yet there is a dearth of validated survey measures for capturing digital literacy of populations with limited prior exposure to technology. In this work, we evaluated and recommend a platform-neutral survey module for measuring digital literacy. Our proposed survey module consists of seven items that are most predictive of ground truth digital literacy measured using participant observation.

We find that when the goal is to measure the digital literacy of users of a specific social media platform (e.g., Facebook, Twitter, or WhatsApp), platform-specific measures – such as the FB module – can

[25] In Online Appendix Table 6, we show that while the full 22-item platform-neutral module improves R^2 over the training data, it does not generalize to unseen data any better than its best 7-item counterpart. Such a module is particularly suitable for measuring digital literacy among populations with a significant fraction of individuals with low Internet experiences, such as in developing countries.

[26] A quintile represents 20% of a given set of values. The first quintile (or quintile rank 1) represents the smallest 20% of values, the second quintile (rank 2) includes values from 20%–40%, and so on.

[27] These users attained a DL score between $[0.89, 1]$, i.e., they either completed 8 or 9 out of a total of 9 DL tasks.

Table 9
Column 1 shows the survey questions of the platform-neutral module, which we recommended based on our evaluation. Column 2 shows the corresponding response options for each question.

Recommended survey module: platform-neutral	Response options
1. Are you able to search/google things online?	Yes; No
How familiar are you with the following computer and Internet-related items? Please choose a number between 1 and 5, where 1 represents no understanding and 5 represents full understanding of the item.:	
2. Internet	1–5
3. Browser	1–5
4. PDF	1–5
5. Bookmark	1–5
6. URL	1–5
7. Torrent	1–5

serve as a good proxy for the underlying ground truth digital literacy. They are attractive because the use of platform-specific features can potentially be inferred through users' natural interactions with the platform without the need for taking a survey.

Our work also highlights the importance of considering contextual relevance and the experiences of the population being studied when designing survey modules and participant observation tasks for measuring digital literacy. As new Internet services, tools, and (social media) application features become part of mainstream Internet uses, the evaluation and addition of new items will be necessary.

In studies that aim to find the causal impact of digital literacy on different outcomes, such as technology adoption and use, ability to distinguish between true and false information, trust in information sources, economic and political behaviors, it is important to consider representative samples from the target population of interest. For example, MTurk samples are known to have higher digital literacy (Hargittai and Shaw, 2020) and thus findings derived from such samples may not hold in the general population of interest.

Our results provide evidence that survey measures can be used to identify low digital literacy populations, who may be left out from the digitalization process or more vulnerable to negative externalities of technology (e.g., misinformation) and thus important targets for interventions.

Finally, our study design and the associated insights – such as the completion rate of different participation observation tasks we used which cover a wide spectrum of digital skills needed to access, understand, and effectively find information online – can (i) inform the design of interventions for improving the digital literacy of individuals and (ii) provide insights about customizing digital literacy programs for different subgroups within a population (e.g., across age groups, education levels and income levels).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ayesha Ali reports financial support was provided by Facebook Research (Unrestricted Gift, Facebook Integrity Foundational Research Award 2020).

Data availability

Data will be made available on request.

Acknowledgments

We thank Sheikh Arsalan-ul-Haque, Saba Rehman, Hijab Tahir, and Zoha Khan for leading the field implementation and data collection. The study was approved by the IRB at Lahore University of Management Sciences (Protocol Number: LUMS-IRB/081921/IAQ-FWA-00019408).

Funding

This work was partially supported by the Facebook Integrity Foundational Research Award 2020 from Facebook Research in the form of an unrestricted gift.

Appendix A

A.1. Descriptive statistics

See Table A.1.

A.2. Efficiency analysis

Next, we conducted the *efficiency* analysis of PO DL tasks, i.e., the time it takes for participants to complete tasks. Our hypothesis is that higher DL users are likely to complete a DL task sooner than low DL users.

Due to different task completion rates, respondent-level completion times (e.g., sum of the completion times of all tasks) can be biased due to selection effects because we only obtain completion times for tasks that are *successfully* completed; see Online Appendix Table 5. As a result, only high DL users are represented in tasks with low completion rates, who may be more efficient than an average Internet user. In such cases, it is useful to examine *intra-task* time variations, i.e., among only those who completed the task successfully. If there exists variations in DL within the sample, we would expect efficiency to be better among higher DL users for *each* task.

Thus, we regress the DL scores on the time taken by participants to complete each DL task, separately. Fig. A.1 shows the horizontal plot of regression coefficients for each DL tasks. We make the following observations:

- All regression coefficients are in the expected direction, i.e., they are negative. This shows that higher completion times are associated with lower DL scores.
- For more difficult tasks – ones with lower completion rates (e.g., finding the birthplace of Quaid-e-Azam) – coefficients are smaller indicating greater homogeneity in the DL scores who completed these tasks.

A.3. Study procedure

At the start of the study, each participant completed a baseline survey. As part of this survey, they first answered a set of questions about their demographics and social media use. Participants then completed the basic Digital Literacy, Facebook, and WhatsApp survey modules. Finally, participants completed a set of participant observation tasks for the ground truth measurement of digital literacy.

For each PO task, the enumerator recorded whether the participant was able to successfully complete the task (coded as 1 if the task was

Table A.1
Descriptive statistics for the sample.

Variable	N	Mean	St. Dev	Min	Median	Max
Age	143	27.80	7.45	18	26	55
Females	143	0.22	0.41	0	0	1
Education level	143	0.98	0.75	0	1	2
Employment status	143	1.68	0.67	0	2	2
Monthly household expenditure (PKR. '000s)	89	51.80	26.15	15.00	40.00	90.00
Users with a WhatsApp account	143	1.00	0.00	0	1	1
Users with a Facebook account	125	0.87	0.33	0	1	1
Users with a YouTube account	134	0.94	0.24	0	1	1
Users with a Twitter account	42	0.29	0.46	0	1	1
Users with a Instagram account	66	0.46	0.50	0	1	1

Notes: Educational level was coded as 0 (if the attained education was below grade 6), 1 (if between grade 6 and 12) and 2 (if above grade 12). Employment status was code as 0 (for unemployed persons or students), 1 (for part-time employed persons), and 2 (full-time employed persons). For household expenditure, the response options included: Less than 10,000; 10,000–20,000; 20,000–30,000; 30,000–50,000; 50,000–70,000; 70,000–90,000; >90,000. Female is a variable equal to 1 if the respondent was a female, 0 for male, and 2 for other. Except age (which was a quantitative variable with no pre-specified category), the other remaining variables were binary (1 if the respondent had an account on the specific social media platform, 0 otherwise).

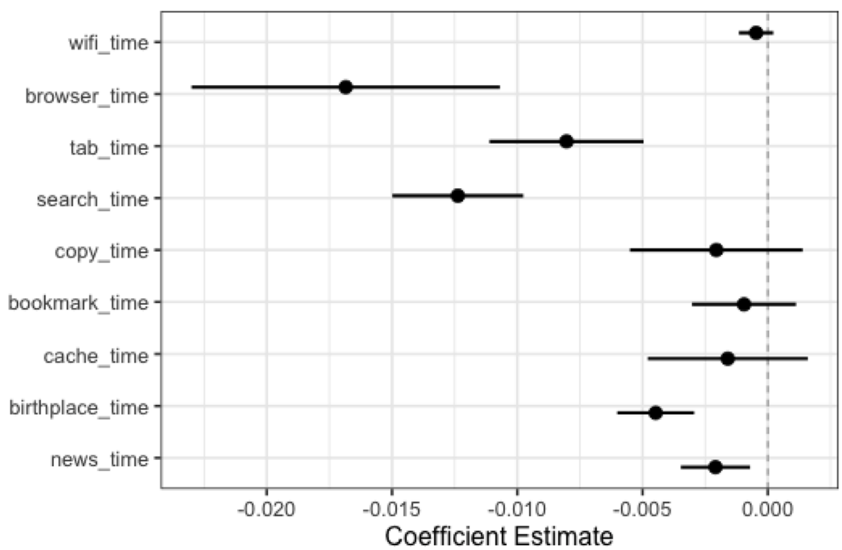

Fig. A.1. Results of OLS regressions. For each DL task, we regress the DL score of respondents on the time taken by participants to complete the task. The plot shows regression coefficients along with 90% confidence intervals.

successfully completed, 0 otherwise) and the time taken to perform the task (recorded via a digital timer). For consistency in measurements and sample size considerations, the ordering of the PO tasks was kept the same across all participants; see Table A.2. Following the PO tasks, the participants took the Terms survey.

While we do not measure the effect of changing the order of survey questions or the PO tasks, it is possible that the order may have had an effect on survey responses and task completions. For example, participants may have become tired or started losing interest towards the end of the survey due to which their responses might have been affected thereby potentially biasing the results.

Finally, we used Android smartphones of the same model for all participants to ensure consistency in measurements. However, participants with different Android phones (e.g., phones with different Android versions) or iPhone users may have lacked familiarity with the study device, which might have impacted their effectiveness as well as efficiency in completing PO tasks.

A.4. Distribution of digital literacy

We now apply two digital literacy measures (Basic DL survey and the Best 7-item platform-specific survey) to determine the distribution of digital literacy on a sample of low- and middle-income social media users in Lahore, Pakistan. The sample was recruited as part of a previous randomized control trial (RCT) in 2019 (Ali and Qazi, 2021). The Terms survey was not conducted on this sample, therefore, we exclude survey measures that include items from the Terms survey. We compare the digital literacy scores predicted by two survey instruments on a subsample of social media ($N = 618$) users who used *both* Facebook and WhatsApp and thus responded to both the surveys that we use for predicting the digital literacy; see Fig. A.2. First, we find that both modules are able to detect significant variations in the digital literacy of respondents. Second, because the Basic DL survey is better able to separate users at the low end of digital literacy, we find greater variation in the predicted digital literacy scores at the lower end. On the other hand, the Best 7-item platform-specific survey is better able to separate users at the high end.[28]

A.5. Pearson's and Spearman's correlation coefficient analysis

In Table A.3, we report the Pearson's and Spearman's Correlation Coefficient along with the corresponding p-values.

[28] In Online Appendix Table 2, we show that combining the Basic DL survey with either the FB survey or the WB survey can result in better prediction of digital literacy.

Table A.2
The order in which the participants completed the digital literacy tasks for ground truth measurement using in-person observations.

Digital Literacy Tasks
1. Connect to a WiFi network
2. Open a mobile browser
3. Open a new tab in the browser
4. Search the term "LUMS" on google.com
5. Copy the URL that appears after searching "LUMS"
6. Bookmark a webpage
7. Clear all cache and cookies from your browser
8. Look up the birthplace of "Quaid-e-Azam"
9. Observe a news headline on social media and find information relevant to the headline using a search engine. (*participants were shown a screenshot of the news headline and asked to find information relevant to the content in the screenshot*)
WhatsApp Tasks
1. Send a chat message (on a given number)
2. Send a voice note
3. Block a contact on WhatsApp
4. Mute notifications for a chat
Facebook Tasks
1. Like/React to a post
2. Comment on a post
3. Search for LUMS facebook page
4. Follow LUMS facebook page
5. Change the privacy settings of a post on your timeline to 'Only Me'

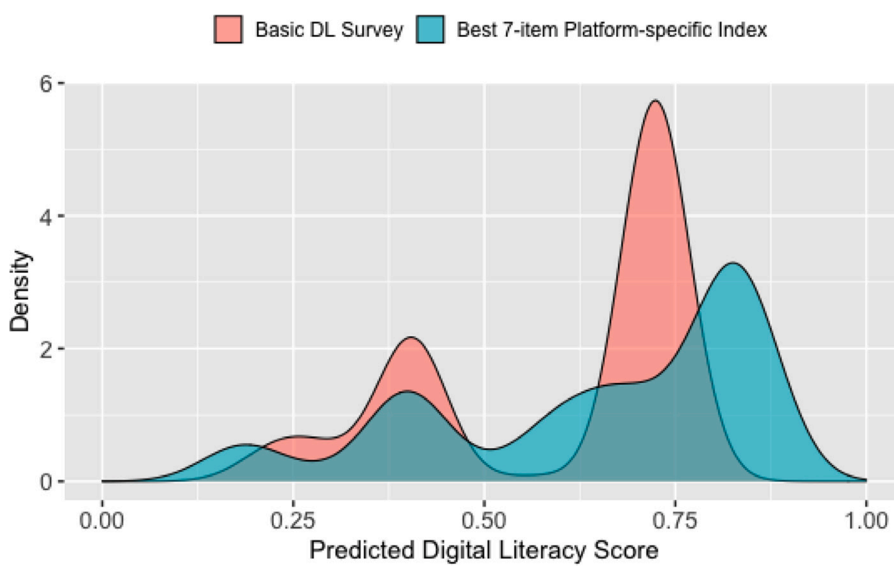

Fig. A.2. Density of predicted digital scores. The plot shows scores predicted by the basic DL survey and the best 7-item platform-specific module on a sample of 618 Facebook users in Lahore, Pakistan.

Table A.3
Correlation between the mean 9-item DL scores and scores obtained via different survey modules. $^{*}p < .10$;$^{**}p < .05$;$^{***}p < .01$.

Survey Instrument	Pearson's Corr. Coeff.	Spearman's Corr. Coef.
1. Basic DL Survey	0.71^{***}	0.70^{***}
2. Internet-related Terms Survey	0.86^{***}	0.87^{***}
3. Knowledge of Facebook Features Survey	0.75^{***}	0.74^{***}
4. Knowledge of WhatsApp Features Survey	0.64^{***}	0.64^{***}
Best 7-item Survey Modules		
α. Best 7-item Internet-related Terms Module	0.86^{***}	0.87^{***}
β. Best 7-item Platform-neutral Module	0.88^{***}	0.87^{***}
γ. Best 7-item Platform-specific Module	0.79^{***}	0.77^{***}
Demographic Characteristics		
a. Education	0.72^{***}	0.71^{***}
b. Income	0.62^{***}	0.63^{***}
c. Age	-0.37^{***}	-0.29^{***}
d. Employment Status	-0.46^{***}	-0.51^{***}
e. Gender	-0.03	0.01

References

Aker, J., Blumenstock, J., 2014. The economic impacts of new technologies in Africa. In: The Oxford Handbook of Africa and Economics. pp. 354–371.

Ali, A., Qazi, I.A., 2021. Countering misinformation on social media through educational interventions: Evidence from a randomized experiment in Pakistan. arXiv:2107.02775.

Ali, A., Qazi, I.A., 2022. Digital literacy and vulnerability to misinformation: Evidence from Facebook users in Pakistan. J. Quant. Descr.: Digit. Media 2, 1–38.

Altmann, A., Tolosi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinform. 26 (10), 1340–1347.

Antin, J., Shaw, A., 2012. Social desirability bias and self-reports of motivation: A study of Amazon mechanical turk in the US and India. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '12, Association for Computing Machinery, New York, NY, USA, pp. 2925–2934.

Archer, E., 2016. rfPermute: estimate permutation p-values for random forest importance metrics. 1, (2), R Package Version.

Biau, G., Scornet, E., 2016. A random forest guided tour. TEST: An Off. J. Span. Soci. Statist. Oper. Res. 25 (2), 197–227.

Brashier, N.M., Schacter, D.L., 2020. Aging in an era of fake news. Curr. Dir. Psychol. Sci. 29 (3), 316–323, PMID: 32968336.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Chun, N., Tang, H., 2018. Do information and communication technologies empower female workers? Firm-level evidence from Viet Nam. (Working Papers id:12804), eSocialSciences.

Cranz, A., 2022. *The Verge*. There are over 3 billion active Android devices. https://www.theverge.com/2021/5/18/22440813/android-devices-active-number-smartphones-google-2021. (Accessed 2 April 2022).

DataReportal, 2022. Digital 2022 global digital overview. trieved from https://datareportal.com/reports/digital-2022-global-overview-report.

DeVon, H.A., Block, M.E., Moyle-Wright, P., Ernst, D.M., Hayden, S.J., Lazzara, D.J., Savoy, S.M., Kostas-Polston, E., 2007. A psychometric toolbox for testing validity and reliability. J. Nurs. Scholarsh. 39 (2), 155–164.

Dimaggio, P., Hargittai, E., Celeste, C., Shafer, S., 2004. Digital inequality: From unequal access to differentiated use. In: Social Inequality. Russell Sage Foundation, pp. 355–400.

Dodson, L.L., Sterling, S.R., Bennett, J.K., 2013. Minding the gaps: Cultural, technical and gender-based barriers to mobile use in oral-language Berber communities in Morocco. In: Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers - Volume 1. ICTD '13, Association for Computing Machinery, New York, NY, USA, pp. 79–88.

EIU Inclusive Internet Index, 2021. The Inclusive Internet Index. https://theinclusiveinternet.eiu.com/explore/countries/PK/ (accessed Oct. 03, 2021).

Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., Moran, S., 2018. Falling for fake news: Investigating the consumption of news via social media. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, Association for Computing Machinery, New York, NY, USA, pp. 1–10.

Gilster, P., 1997. Digital Literacy. Wiley Computer Pub, New York.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., Lazer, D., 2019. Fake news on Twitter during the 2016 U.S. presidential election. Science 363 (6425), 374–378.

Guess, A.M., Munger, K., 2020. Digital literacy and online political behavior. http://osf.io/3ncmk.

Guess, A.M., Munger, K., 2022. Digital literacy and online political behavior. Political Sci. Res. Methods 1–19.

Guriev, S., Melnikov, N., Zhuravskaya, E., 2020. 3G Internet and Confidence in Government*. Q. J. Econ. 136 (4), 2533–2613.

Hargittai, E., 2005. Survey measures of web-oriented digital literacy. Soc. Sci. Comput. Rev. 23 (3), 371–379.

Hargittai, E., 2009. An update on survey measures of web-oriented digital literacy. Soc. Sci. Comput. Rev. 27 (1), 130–137.

Hargittai, E., Hsieh, Y.P., 2012. Succinct survey measures of web-use skills. Soc. Sci. Comput. Rev. 30 (1), 95–107.

Hargittai, E., Micheli, M., 2019. Internet skills and why they matter. In: Graham, M., Dutton, W.H. (Eds.), Society and the Internet : How Networks of Information and Communication are Changing Our Lives. Oxford University Press, Oxford, pp. 109–124.

Hargittai, E., Piper, A., Morris, M., 2018. From internet access to internet skills: digital inequality among older adults. Univers. Access in the Inf. Soc. 1–10.

Hargittai, E., Shaw, A., 2020. Comparing internet experiences and prosociality in amazon mechanical turk and population-based survey samples. Socius 6, 2378023119889834.

Hjort, J., Tian, L., 2021. The economic impact of internet connectivity in developing countries. INSEAD Working Paper 2021/68/EPS.

ITU, 2021. 2.9 billion people still offline. New data from ITU suggest 'COVID connectivity boost' – but world's poorest being left far behind. https://www.itu.int/en/mediacentre/Pages/PR-2021-11-29-FactsFigures.aspx.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., Westwood, S.J., 2019. The origins and consequences of affective polarization in the United States. Annu. Rev. Political Sci. 22 (1), 129–146.

Jack, W., Suri, T., 2014. Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution. Amer. Econ. Rev. 104 (1), 183–223.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R. Springer.

Jayachandran, S., Biradavolu, M., Cooper, J., 2021. Using Machine Learning and Qualitative Interviews to Design a Five-Question Women's Agency Index. IZA Discussion Papers 14221, Institute of Labor Economics (IZA).

Lee, J., Devanarayan, V., Barrett, Y.C., Weiner, R., Allinson, J., Fountain, S., Keller, S., Weinryb, I., Green, M., Duan, L., Rogers, J., Millham, R., O'Brien, P., Sailstad, J., Khan, M., Ray, C., Wagner, J., 2006. Fit-for-purpose method development and validation for successful biomarker measurement. Pharm. Res. 23, 312–328.

Levy, R., 2021. Social media, news consumption, and polarization: Evidence from a field experiment. Amer. Econ. Rev. 111 (3), 831–870.

Medhi Thies, I., 2015. User interface design for low-literate and novice users: Past, present and future. Found. Trends Hum.-Comput. Interact. 8 (1), 1–72.

Muda, R., Pennycook, G., Pieńkosz, D., Bialek, M., 2021. People are worse at detecting fake news in their foreign language. OSF Preprints.

Najam, Bari, 2018. Pakistan national human development report: Unleashin the potential of a Young Pakistan. https://www.pk.undp.org/content/pakistan/en/home/library/human-development-reports/PKNHDR.html.

Osborne, J., Blanchard, M., 2011. Random responding from participants is a threat to the validity of social science research results. Front. Psychol. 1.

Qazi, I.A., Qazi, Z.A., Ali, A., Abdullah, M., Habib, R., 2021. Rethinking web for affordability and inclusion. In: Proceedings of the Twentieth ACM Workshop on Hot Topics in Networks. HotNets '21, Association for Computing Machinery, New York, NY, USA, pp. 9–15.

Rains, S.A., Tsetsi, E., 2017. Social support and digital inequality: Does internet use magnify or mitigate traditional inequities in support availability? Commun. Monogr. 84 (1), 54–74.

Sirlin, N., Epstein, Z., Arechar, A.A., Rand, D.G., 2021. Digital literacy is associated with more discerning accuracy judgments but not sharing intentions. Harvard Kennedy School (HKS) Misinformation Review.

Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8, 25.

Strobl, C., Zeileis, A., 2008. Danger: High power! – exploring the statistical properties of a test for random forest variable importance. In: COMPSTAT 2008 - Proceedings in Computational Statistics.

van Deursen, A., van Dijk, J., 2009. Improving digital skills for the use of online public information and services. Gov. Inf. Q. 26 (2), 333–340, Building the Next-Generation Digital Government Infrastructures.

Vashistha, A., Saif, U., Raza, A.A., 2019. The internet of the orals. Commun. ACM 62 (11), 100–103.

Wheeler, L., Garlick, R., Johnson, E., Shaw, P., Gargano, M., 2022. LinkedIn(to) job opportunities: Experimental evidence from job readiness training. Am. Econ. J.: Appl. Econo. forthcoming.

World Bank, 2021. World Development Report 2021: Data for Better Lives. https://www.worldbank.org/en/publication/wdr2021.

Zhuravskaya, E., Petrova, M., Enikolopov, R., 2020. Political effects of the internet and social media. Annu. Rev. Econ. 12 (1), 415–438.

Zillien, N., Hargittai, E., 2009. Digital distinction: Status-specific types of internet usage. Soc. Sci. Q. 90 (2), 274–291.