Online Appendix for "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* [b] *Department of Computer Science*
[a,b] *Lahore University of Management Sciences, DHA, Lahore Cantt, 54792, Lahore, Pakistan*

Contents

A Analysis of 4-item Information Retrieval Tasks

We also conducted analysis for a different ground truth measure of digital literacy comprising just the four information retrieval tasks. These include Tasks 4, 6, 8, and 9 in Table **??**. Table 1 shows the completion ratio for each task. Observe that these tasks also cover a wide range of completion rates similar to the 9-item DL score. However, the 9-item score is able to detect more *granular* differences in digital literacy. Table 2 shows the correlation of different survey modules with the ground truth DL score considering this 4-item construct. We find these correlations to be similar in magnitude compared to the 9-item DL score.

B Demographic Survey and Demographic Correlates

Table 3 shows that questions, response options, and the coding we used for the demographics survey. Table 4 presents the results of a OLS regression that aims to predict the DL score using age, income, education, gender, and employment as independent variables.

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

Digital Literacy Task (*information retrieval related*)	**Completion Rate**
1. Search the term "LUMS" on google.com	0.82
2. Look up the birthplace of "Quaid-e-Azam"	0.73
3. 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
4. Copy the URL that appears after searching "LUMS"	0.45

Table 1 Four Digital Literacy Tasks related to Information Retrieval and their completion rates obtained using In-Person Observations. $N = 143$)

Survey Instrument	**Pearson's Corr. Coeff.**	**Spearman's Corr. Coef.**
1. Basic DL Survey	0.70^{***}	0.68^{***}
2. Internet-related Terms Survey	0.82^{***}	0.83^{***}
3. Knowledge of Facebook Features Survey	0.70^{***}	0.70^{***}
4. Knowledge of WhatsApp Features Survey	0.59^{***}	0.59^{***}
Best 7-item Survey Modules		
α. Best 7-item Internet-related Terms Module	0.82^{***}	0.82^{***}
β. Best 7-item Platform-neutral Module	0.83^{***}	0.83^{***}
γ. Best 7-item Platform-specific Module	0.73^{***}	0.72^{***}
Demographic Characteristics		
a. Education	0.71^{***}	0.69^{***}
b. Income	0.57^{***}	0.58^{***}
c. Age	-0.33^{***}	-0.25^{***}
d. Employment Status	-0.40^{***}	-0.45^{***}
e. Gender	-0.04	0.01

Table 2 Correlation between the mean of the 4-item Information Literacy score (obtained via participant observation) and scores obtained via different survey modules. The four items include Tasks 4, 6, 8, and 9 in Table 1 in the paper. $^{*}p < .10;^{**}p < .05;^{***}p < .01$.

C Task Completion Time Analysis

Table 5 shows the mean completion time for the nine tasks we considered.

D Extended Set of Results

Table 6 shows the performance results for the various survey modules we evaluated in terms of MSE and R^2. This extended set of results also include different combinations of survey modules (e.g., combining Basic DL survey and the FB survey). These results suggest that combining the Basic DL survey with either the FB survey or the WA survey leads to better prediction of DL scores compared to using these surveys individually.

E Best 7-item Survey Instruments

Table 7 shows 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). Table 8 shows the feature importances of each survey item along with their corresponding p-values.

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 rfPer-

Survey Question	**Response Options**	**Coding**	**Median**
1. What is your age?	*Direct input from user (no pre-specified options)*	None	26 years
2. What is your highest level of education?	*None; Primary School (till Grade 6); Middle School (from Grade 6 to Grade 8); Metric; Intermediate; Bachelors; Masters; Above Masters; Diploma*	0 (below Grade 6); 1 (between Grade 6 and 12); 2 (above Grade 12)	1 (between Grade 6 and 12)
3. What is your current status of employment?	*Full-time employed; Part-time employed; Self-employed; Unemployed; Student*	0 (Unemployed or Student); 1 (Part-time employed); 2 (full-time employed or self-employed)	2
4. What is aggregate expenditure of your household per month?	*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; Don't know; Don't want to say*	None	PKR 40,000
5. What is your gender?	*Female; Male; Other*	0 (Male), 1 (Female), 2 (Other)	

Table 3 The table reports variable mean and standard deviations.

mute package in R. For each survey model, 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.

Table 4 Demographic Correlates of Digital Literacy

	Dependent variable:
	DL Score
Age	−0.079
	(0.069)
Income	0.245**
	(0.099)
Education	0.532***
	(0.081)
Employment	−0.208**
	(0.080)
Gender	−0.131*
	(0.072)
Constant	−0.000
	(0.065)
Observations	89
R^2	0.649
Adjusted R^2	0.627
Residual Std. Error	0.610 (df = 83)
F Statistic	30.647*** (df = 5; 83)
Note:	*p<0.1; **p<0.05; ***p<0.01

Digital Literacy Task	**Completion Rate**	**Mean Completion Time (secs)**
1. Connect to a WiFi network	0.87	13.6
2. Search the term "LUMS" on google.com	0.82	11.4
3. Open a mobile browser	0.78	6.3
4. Look up the birthplace of "Quaid-e-Azam"	0.73	22.2
5. Open a new tab in the browser	0.52	7.5
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	36.3
7. Copy the URL that appears after searching "LUMS"	0.45	10.1
8. Bookmark a webpage	0.34	14.4
9. Clear all cache and cookies from your browser	0.30	17.1

Table 5 Mean completion time for all Digital Literacy Tasks using In-Person Observations. $N = 143$.

Survey Modules	**Items**	***N***	**DL**	**MSE**	R^2 **(Train)**	R^2 **(OOB)**
1. Basic DL Survey	7	143	0.59	0.048	0.60	0.52
2. Internet-related Terms Survey	15	143	0.59	**0.022**	**0.94**	**0.78**
3. Knowledge of Facebook Features Survey	9	125	0.63	0.039	0.72	0.56
4. Knowledge of WhatsApp Features Survey	9	143	0.59	0.056	0.55	0.45
Composite Survey Modules						
1. Platform-neutral Survey	22	143	0.59	0.020	0.95	0.80
2. Platform-specific Survey	18	125	0.63	0.038	0.82	0.58
3. Basic DL + FB Survey	16	125	0.63	0.022	0.94	0.76
4. Basic DL + WA Survey	16	143	0.59	0.021	0.95	0.79
Platform-specific PO Tasks & Demographics Survey Module						
1. Facebook Tasks	5	34	0.53	0.038	0.76	0.66
2. WhatsApp Tasks	4	44	0.53	0.055	0.63	0.54
3. Demographics Survey	5	89	0.62	0.035	0.75	0.59
Best 7-item Survey Module						
α. Best 7-item Internet-related Terms survey	7	143	0.59	0.022	0.89	0.78
β. Best 7-item Platform-neutral survey	7	143	0.59	0.019	0.90	0.80
γ. Best 7-item Platform-specific survey	7	125	0.63	0.033	0.73	0.63

Table 6 All results table. Mean Squared Error and R^2 for different survey instruments including individual, composite, and best 7-item surveys on the training and OOB samples using a Random Forest Regression Model.

Question/Feature	Survey Type	Term	Global	Plat-Neutral	Plat-Specific
1. Internet	Terms	Internet	Internet	Internet	
2. MP4	Terms				
3. Browser	Terms	Browser	Browser	Browser	
4. Search engine	Terms				
5. MB/GB	Terms				
6. PDF	Terms	PDF	PDF	PDF	
7. Bookmark	Terms	Bookmark	Bookmark	Bookmark	
8. JPG	Terms	JPG			
9. URL	Terms	URL	URL	URL	
10. Cookies	Terms				
11. Torrent	Terms	Torrent	Torrent	Torrent	
12. Podcasting	Terms				
13. Malware	Terms				
14. Phix (bogus)	Terms				
15. Jcrypt (bogus)	Terms				
16. wifi_mobile	Basic DL				
17. search	Basic DL		search	search	
18. read_text	Basic DL				
19. without_assistance	Basic DL				
20. viewing	Basic DL				
21. creating	Basic DL				
22. sharing	Basic DL				
23. fb_create_post	FB				
24. fb_like_post	FB				
25. fb_comment	FB				fb_comment
26. fb_privacy_account	FB				fb_privacy_account
27. fb_privacy_post	FB				fb_privacy_post
28. fb_report_account	FB				
29. fb_hide_ad	FB				fb_hide_ad
30. fb_hide_post	FB				
31. fb_sponsored	FB				fb_sponsored
32. wa_view_chat	WA				
33. wa_reply_chat	WA				
34. wa_voice_note	WA				
35. wa_forward_message	WA				
36. wa_delete_message	WA				
37. wa_report_user	WA				wa_report_user
38. wa_block_user	WA				
39. wa_message_seen	WA				wa_message_seen
40. wa_audio_speed	WA				
MSE		0.022	0.019	0.019	0.033
R^2 (OOB Data)		0.78	0.80	0.80	0.63

Table 7 Best 7-item Survey instruments for different categories.

Question/Feature	Terms Survey	Basic DL Survey	FB Survey	WA Survey
1. PDF	(320.1***, 3.68***)			
2. Browser	(261.4***, 0.96**)			
3. Internet	(261.4**, 1.33**)			
4. Bookmark	(213.6**, 1.85***)			
5. Cookies	(175.8**, 0.33)			
6. Search engine	(160.9*, 0.89)			
7. MB/GB	(156.2, 0.34)			
8. URL	(151.8*, 0.94**)			
9. Podcasting	(146.9, 0.29)			
10. Torrent	(142.1**, 0.63**)			
11. JPG	(137.2**, 1.32***)			
12. MP4	(125.1, 0.47)			
13. Malware	(115.1, 0.24)			
14. Phix (bogus)	(-21.5, 0.07)			
15. Jcrypt (bogus)	(-26.4, 0.07)			
16. search		(418.6***, 2.89***)		
17. read_text		(392.6***, 1.92***)		
18. wifi_mobile		(271.1***, 1.89***)		
19. without_assistance		(209.1**, 0.49*)		
20. creating		(44.1, 0.25)		
21. viewing		(0.0, 0.02)		
22. sharing		(-24.7, 0.19)		
23. fb_privacy_account			(322.3***, 2.1***)	
24. fb_sponsored			(267.3**, 1.2***)	
25. fb_privacy_post			(260.5**, 2.0***)	
26. fb_hide_ad			(232.9**, 1.11***)	
27. fb_comment			(166.2*, 0.45*)	
28. fb_hide_post			(93.2, 0.54*)	
31. fb_create_post			(35.0, 0.28)	
29. fb_report_post			(18.0, 0.25)	
30. fb_like_post			(0.0, 0.14***)	
32. wa_report_user				(414.6***, 2.55***)
33. wa_message_seen				(354.4***, 1.69***)
34. wa_block_user				(267.0***, 1.56***)
35. wa_reply_chat				(170.0, 0.59***)
36. wa_forward_message				(73.6, 0.45***)
37. wa_audio_speed				(56.2, 0.37)
38. wa_voice_note				(6.3, 0.16)
39. wa_view_chat				(-0.92, 0.14)
40. wa_delete_message				(-38.2, 0.12)
MSE	0.017	0.036	0.029	0.042
R^2 (OOB Samples)	0.535	0.599	0.577	0.451

Table 8 (%IncMSE, IncNodePurity) feature importance values across different surveys. $^{*}p < .10;^{**}p < .05;^{***}p < .01$.