When Scales Fail to Measure Up: How Not to Measure Social Media Privacy–Findings of a Representative Survey in 16 Countries

Habiba Farzand
University of Glasgow
Glasgow, United Kingdom
University of Bristol
Bristol, United Kingdom
habiba.farzand@gmail.com

Ali Farooq
Computer and Information Sciences
University of Strathclyde
Glasgow, United Kingdom
ali.farooq@strath.ac.uk

Joni Salminen
University of Vaasa
Vaasa, Finland
jonisalm@uwasa.fi

Bernard J Jansen
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
jjansen@acm.org

Abstract

Privacy scales are scientific instruments measuring constructs, such as users' concerns, preferences, or behaviour; however, the development and validation of privacy scales are often limited to populations in the Global North. This poses a challenge for evaluating and generalising the scale results in the Global South. We present the results of a replication study on the external validity of the Privacy-Defensive Behaviour Scale (PDBS) with a representative sample of 8140 participants from 16 Middle East and North Africa (MENA) countries. Our results show that the PDBS, when administered in the MENA region, has a poor model fit, does not support the factor structure of the original scale, and exhibits low reliability and validity. Our results highlight the need for cross-cultural scale validation and adaptation. We discuss possible underlying reasons and best practices for scale validation studies.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**; **Empirical studies in HCI**;

Keywords

Privacy Defensive Behavior, Scale Validity, Cross-Cultural Study, MENA

ACM Reference Format:
Habiba Farzand, Ali Farooq, Joni Salminen, and Bernard J Jansen. 2025. When Scales Fail to Measure Up: How Not to Measure Social Media Privacy–Findings of a Representative Survey in 16 Countries. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3706599.3720138

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3720138

1 Introduction

People, firms, and governments around the world use social media for maintaining social relationships, sharing information, and doing businesses [9, 19, 26]. According to an estimate, there are 5.85 billion social media users worldwide [53]. The pervasiveness of social media has spurred debates on privacy and its impact on users' values, autonomy [29], freedom [52], rights [4] and the diminishing demarcation of public-private spheres [45]. Privacy issues are intricately connected with the design of social media platforms, where users are encouraged to share their personal information [34, 51]. Studies have shown that users have multiple privacy concerns related to service providers, other users, and regulators [21, 24, 33], which can result in social media discontinuation [20].

In privacy research, researchers have proposed practices for individual privacy management [3], collaborative privacy management [32], collective privacy management [49], and interpersonal privacy management [37]. In addition to these, several privacy scales such as Internet Users Information Privacy Concerns [42], Value for Other People's Privacy (VOPP) [27], and Out-of-Device Privacy Scale (ODPS) [23] have been proposed to measure users' privacy attitudes and needs. However, one of the limitations of existing privacy methods research is that they are mainly based on samples from a single country. For example, VOPP was developed using only a US sample, and ODPS was developed using only a UK sample. Cross-country investigations of the reliability and validity of methods are crucial, as privacy is deeply rooted in culture [1]. Existing research on cross-cultural studies has shed light by comparing the level of privacy concerns and privacy behaviours in different countries [14, 15, 36, 39, 47] and revealed that privacy conceptualisation developed in one country might not hold in another. Therefore, examining their reliability and validity in culturally different contexts is necessary to make privacy constructs reliable and valid across different cultures.

In the context of scales, *reliability* refers to the degree to which the survey items (i.e., questions, statements) are consistent with each other [12]. In turn, *validity* means that the construct accurately measures the construct it aims to measure [11]. In this research, we investigate the reliability and validity of the Privacy Defensive Behaviour Scale (PDBS) [3], originally developed using a sample (N=335) from Australia. We investigate the PDBS using a representative sample from sixteen countries in the MENA region (N=8140) in the social media context.

For this, we address the following research question: *How is the reliability and validity of a social media privacy defensive behaviour scale impacted when administered in the MENA cultures?*

More specifically, we analyse the underlying dimensions of the PDBS [3] using confirmatory factor analysis to verify the model fit, reliability, and validity of PDBS. Our results show that PDBS had poor model fit and weak reliability *in all 16 countries* in the MENA region. We then re-examined the item loadings and removed the items with low loadings to improve the model fit indices. However, the results still showed poor model fit and weak reliability.

Overall, this research makes the following contributions. First, to emphasise the importance of cross-cultural scale validation studies, we contribute a reliability and validity assessment study of the PDBS with a representative sample of 8140 participants from the MENA region. Second, to assist researchers, we reflect on the results and discuss the possible underlying reasons for the scales' low validity when administered to different populations.

2 Related Work

2.1 Challenges in Scale Development and Validation

Numerous privacy scales exist to capture granular concepts that fall under the umbrella of privacy, such as information privacy concerns [42], value for other people's privacy [27], and out-of-device privacy [23]. While similar constructs capture essential information, their development and validation come with many challenges. Colnago et al. [16] examined the interpretation of the items on four commonly used privacy scales: Internet Users' Information Privacy Concern (IUIPC), Westin's Privacy Segmentation Index, Concern for Information Privacy (CFIP), and Global Information Privacy Concern (GIPC), with a representative sample from the United States. The researchers found a significant misalignment between what the construct aimed to measure and what participants perceived it was measuring. One of the underlying reasons was the ambiguity of natural language. Although the investigated scales have been rigorously developed and tested in English-speaking countries, they were still perceived differently by people from other English-speaking countries. Many scales have translations available to overcome language discrepancies, such as the User Experience Questionnaire (UEQ) [38]. However, translations into other languages for privacy scales are still missing.

Differences in time could also indicate variations in the results. An example of the impact of time on scale validation is the study by Gross [25] that investigated the validity of the Internet Users Information Privacy Concern Scale (IUIPC), originally developed as a 10-item scale by Malhotra et al. [42] in 2004 and provided evidence of an eight-item IUIPC scale which offered improved reliability and validity in contract to Malhotra's 10-item IUIPC scale [25].

Furthermore, existing research investigating cross-cultural differences in measurements (for example, [14, 39] considers a couple (two or three) countries to compare and uses non-representative samples. To sum up, time, culture and language interpretation plays a vital role in the validity of scales and, therefore, require validation studies over time with different populations and representative samples to ensure accurate measurements.

2.2 Motivation for Validity Assessment

Our research goal is in line with the concept of conceptual replication [7, 17, 50]. Conceptual replication is a study that investigates the same research question as the original study but with controlled changes in participants, methods, or settings. Conceptual replication brings the advantage of generalizability, providing external validity [10]. Conceptual replication studies have been recognised as a significant part of scientific literature as they assist in broadening the understanding of the subject [17]. Related work favours conceptual replication over direct replication as it extends the results and provides more vigorous evidence-based results [50]. Therefore, in this paper, we verify the external validity of the PDBS [3] by utilising the conceptual replication approach. The PDBS was introduced in the original power-responsibility equilibrium theory (PRE) framework [40] and carried forward in the extended PRE framework [3]. The item loadings, composite reliability, and average variance extracted from the original study of the Privacy Defensive Behaviour Scale are presented in Table A1 in the Appendix.

3 Methodology

3.1 Data Collection

An online survey was conducted in the 16 MENA countries of Algeria, Bahrain, Egypt, Iraq, Jordan, KSA, Kuwait, Lebanon, Libya, Morocco, Oman, Palestine, Qatar, Tunisia, UAE, and Yemen to collect a population-representative dataset in terms of gender and age and, where possible, education through a survey firm TGM[1]. These countries were selected due to similar cultures and languages; at least one official language of all these countries is Arabic. A couple of countries in the region were excluded due to volatile situations, for example, Syria, and heterogeneous cultural and political landscapes, for example, Iran and Israel. TGM has registered panellists from many countries recruited through a multi-stage stratified sampling technique to ensure population representativeness. Random sampling was employed at each stage. TGM pays panellists depending on the length and type of the survey. In our case, each participant was compensated above the minimum hourly wage in the US. The ethics committee at one of the author's home institutions approved the study.

3.2 Questionnaire Design

The questionnaire started with an introduction detailing the purpose of the study, the nature of participation, and information about data confidentiality, followed by an informed consent form. Only those participants who gave their consent proceeded with the study. The questionnaire began with general questions about internet and social media use, regarding how many hours spent, followed by privacy defensive behaviour scale statements as shown in Table A1.

[1]https://tgmresearch.com

The PDBS scale was adapted to the social media context and deployed on a 5-point Likert scale: responses ranging from strongly disagree (1) to strongly agree (5) with a neutral middle anchor point neither disagree nor agree (3). Statements were presented in random order to minimize order bias. Two attention check questions were included in the survey in random places, whereas TGM applied additional quality checks. The questionnaire went through internal checks in which experts in HCI and privacy examined the statements. The questionnaire was then translated into Arabic using the back-translation method [8]. Minor changes in wording were implemented before data collection.

3.3 Pilot Testing

To ensure the content and face validity of the questionnaire, a pilot test (n=100) was conducted among social media users of one of the 16 countries to evaluate item clarity, relevance, and comprehensibility. After making the necessary adjustments, especially in the Arabic language questionnaire, a full-scale data collection was launched in 16 countries in the summer of 2023 in English and Arabic. Most of the respondents preferred to respond in Arabic (75.7%). No personally identifiable information was collected.

3.4 Participants

A sample size of 500 for each country was determined to sufficiently present the target population of the country with a reasonable error margin (+/-4 percentage points), calculated $1/\sqrt{n}$) [2]. We recruited a representative sample of 8140 participants from the 16 MENA countries while aiming for at least 500 participants from each country. However, in some countries, we could only recruit slightly fewer than 500 participants despite efforts. This sample size limitation should be considered when using the results of this study in future work. Table A2 in the Appendix presents the demographic details.

3.5 Step 1: Model Assessment

Initial Data Analysis: Before proceeding with the model assessment of the PDBS, we checked for the suitability of the dataset to perform confirmatory factor analysis by calculating the Kaiser-Meyer-Olkin (KMO) Measures of Sampling Adequacy (MSA) test [13]. A KMO value of 0.60 or higher is appropriate for factor analysis. In our case, the dataset for all 16 countries passed the KMO MSA test along with Bartlett's test of Sphericity and hence was considered suitable for data analysis. Full details can be found in Appendix.

Confirmatory Factor Analysis: Confirmatory factor analysis is a model-testing approach to test and verify the efficacy of measurement models. We performed a confirmatory factor analysis on the data collected for the 16 countries. We reported the following fit indices: the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI) [6]. For a model to indicate exceptional fit, it must meet the following criteria of the fit indices: CFI > 0.90, TLI >0.90, and RMSEA < 0.08 [5, 35, 41, 44]. The results show that CFI was achieved greater than 0.9 in only two countries, Iraq and Jordan, whereas TLI was less than 0.9 across all countries. RMSEA was acceptable in four countries: Libya, Iraq, Jordan, and Palestine. Overall, the results suggest that the recommended thresholds for confirmatory factor analysis were not met in any country under investigation. Table 1 shows the results of the confirmatory factor analysis.

Country	CFI	TLI	RMSEA
Algeria	0.837	0.619	0.107
Egypt	0.815	0.569	0.117
Libya	0.892	0.747	0.068
Morocco	0.797	0.526	0.122
Tunisia	0.809	0.555	0.119
Iraq	0.921	0.817	0.068
Jordan	0.912	0.854	0.079
Lebanon	0.827	0.595	0.096
Palestine	0.871	0.699	0.078
Bahrain	0.835	0.616	0.091
Kuwait	0.821	0.583	0.11
Oman	0.786	0.5	0.113
Qatar	0.804	0.543	0.129
KSA	0.857	0.667	0.115
UAE	0.816	0.57	0.101
Yemen	0.712	0.328	0.128

Table 1: Results of the Confirmatory Factor Analysis (CFI, TLI, RMSEA) of the 6-item Privacy Defensive Behavior Scale. The cells highlighted with a green background indicate where the threshold for the fit indices was achieved.

Tests of Reliability and Validity: Reliability refers to internal consistency [11, 18]. To assess reliability, we calculated reliability using Cronbach's Alpha. A coefficient of 0.7 or above is considered acceptable. In our case, reliability in all 16 countries was less than 0.7, indicating poor reliability. We then checked for composite reliability (CR), which should be above 0.60 to demonstrate good internal consistency. Good composite reliability was observed in all countries except Libya, Jordan, and Yemen (i.e., < 0.60). We then proceeded to compute the average variance extracted (AVE). The average variance extracted should be above 0.50 to provide evidence of convergence validity. We observed AVE to be less than 0.5 across all countries. Therefore, we can conclude that there is a lack of convergent validity. The results of the reliability and validity evaluation are presented in Table 2.

3.6 Step 2: Additional Model Assessment

In the previous selection, we observed that the 6-item PDBS dataset did not meet the model fit indices' requirements and reliability (Cronbach's Alpha). The threshold of composite reliability was achieved in most cases; however, the average variance extracted (AVE) threshold was not achieved. As the next step, we inspected the results to identify possible reasons. One method of improving model fit indices is to remove items with low loadings. In our dataset, PDBS 1, PDBS 4, and PDBS 5 had low loadings across most of the countries (as shown in Table 2).

Considering the items with poor loadings, we followed a stepwise model selection process to improve the model fit indices. We eliminated one item at a time and reviewed the model fit indices. The results revealed that the model fit indices did not improve in

Country	Item ID	Loadings	Reliability	CR	AVE	Country	Item ID	Loadings	Reliability	CR	AVE
Algeria	PDBS_1	0.608	0.627	0.629	0.237	Palestine	PDBS_1	0.333	0.594	0.605	0.212
	PDBS_2	0.679					PDBS_2	0.467			
	PDBS_3	0.474					PDBS_3	0.642			
	PDBS_4	0.242					PDBS_4	0.399			
	PDBS_5	0.416					PDBS_5	0.503			
	PDBS_6	0.367					PDBS_6	0.345			
Egypt	PDBS_1	0.491	0.661	0.664	0.250	Bahrain	PDBS_1	0.593	0.614	0.610	0.222
	PDBS_2	0.564					PDBS_2	0.67			
	PDBS_3	0.588					PDBS_3	0.43			
	PDBS_4	0.44					PDBS_4	0.348			
	PDBS_5	0.442					PDBS_5	0.274			
	PDBS_6	0.454					PDBS_6	0.385			
Libya	PDBS_1	0.335	0.579	0.590	0.202	Kuwait	PDBS_1	0.396	0.664	0.669	0.256
	PDBS_2	0.465					PDBS_2	0.46			
	PDBS_3	0.624					PDBS_3	0.592			
	PDBS_4	0.304					PDBS_4	0.416			
	PDBS_5	0.432					PDBS_5	0.551			
	PDBS_6	0.463					PDBS_6	0.585			
Morocco	PDBS_1	0.580	0.645	0.640	0.242	Oman	PDBS_1	0.541	0.632	0.633	0.231
	PDBS_2	0.694					PDBS_2	0.633			
	PDBS_3	0.393					PDBS_3	0.439			
	PDBS_4	0.32					PDBS_4	0.317			
	PDBS_5	0.381					PDBS_5	0.444			
	PDBS_6	0.479					PDBS_6	0.448			
Tunisia	PDBS_1	0.371	0.631	0.641	0.235	Qatar	PDBS_1	0.478	0.688	0.692	0.276
	PDBS_2	0.525					PDBS_2	0.55			
	PDBS_3	0.56					PDBS_3	0.659			
	PDBS_4	0.338					PDBS_4	0.45			
	PDBS_5	0.519					PDBS_5	0.486			
	PDBS_6	0.547					PDBS_6	0.502			
Iraq	PDBS_1	0.563	0.634	0.639	0.235	KSA	PDBS_1	0.526	0.691	0.699	0.287
	PDBS_2	0.586					PDBS_2	0.667			
	PDBS_3	0.486					PDBS_3	0.577			
	PDBS_4	0.301					PDBS_4	0.325			
	PDBS_5	0.445					PDBS_5	0.502			
	PDBS_6	0.471					PDBS_6	0.558			
Jordan	PDBS_1	0.410	0.640	0.647	0.241	UAE	PDBS_1	0.443	0.629	0.632	0.223
	PDBS_2	0.585					PDBS_2	0.517			
	PDBS_3	0.522					PDBS_3	0.493			
	PDBS_4	0.306					PDBS_4	0.42			
	PDBS_5	0.514					PDBS_5	0.447			
	PDBS_6	0.552					PDBS_6	0.508			
Lebanon	PDBS_1	0.398	0.621	0.625	0.220	Yemen	PDBS_1	0.447	0.588	0.591	0.201
	PDBS_2	0.547					PDBS_2	0.537			
	PDBS_3	0.497					PDBS_3	0.373			
	PDBS_4	0.381					PDBS_4	0.28			
	PDBS_5	0.478					PDBS_5	0.448			
	PDBS_6	0.493					PDBS_6	0.545			

Table 2: The reliability and validity assessment results of the 6-item PDBS. The cells highlighted with a green background indicate where the threshold for the fit indices was achieved.

most countries after removing all three items individually and remained below the recommended threshold. This shows that despite removing items with low loadings, the model still performs poorly for most countries. Therefore, it can be concluded that the model cannot be recommended to capture privacy-defensive behaviour for users in the MENA region. The appendix presents the results of the model fit indices upon removing each of the three items with low loadings. We also inspected the model by removing all three low-loading items one by one in a few randomly selected countries; however, the model statistics did not improve. We did not remove items when doing so would have left less than three items, as a factor should ideally have at least three items. Next, we hypothesized that a second-order factor structure might be more appropriate and tested this on a subset of countries. However, the resulting model fit indices did not yet match the expected thresholds. This shows that the current model does not work for the population under investigation (i.e., MENA region).

4 Discussion and Future Work Directions

In this study, we explore the reliability and validity of scales when administered in cultural contexts different from the ones in which they were developed. We present a case study of the Privacy Defensive Behaviour Scale and provide evidence that scales might fail to measure underlying constructs in cross-cultural investigations, i.e. scales that work for one population may not accurately capture the same construct for another population. In this section, we reflect on the possible underlying reasons for the low validity of scales when administered to different populations. Based on this, we derive future work directions.

Uneven Representation Among Countries: A study by Linxen et al. shows that 73% of the publications published at the CHI conference are based on population samples from Western countries. Moreover, the skew towards these countries in usable security and privacy research is greater [28]. However, prior research also provides evidence that there is great variability in experimental results across populations [30]. Along the same lines, in this study, we provide evidence that scales developed in Western countries do not accurately measure constructs when administered in different countries. To this end, we advocate for increased geographic diversity and replication studies in non-Western countries.

Sources of Item Generation: Item generation is one of the most crucial steps in scale development after the definition of the construct. There are two ways for item generation: (1) deductive, in which the items are theorized from the literature, and (2) inductive, in which a group of people is asked about their perception on a specific topic [31]. While both approaches have pros and cons, using a mixed approach has been recommended [6, 43]. A mixed approach is also well-suited, as it helps to overcome the limitations of using each method independently. Recent scale development works have been considering following a mixed inductive and deductive approach to formulate the items [23, 27], and this was also seen in the development of PDBS. However, in the case of PDBS, it remains unclear what sources of prior work were selected for investigation and included based on what criteria, perhaps raising questions of its validity. This selection of previous work might have introduced a bias or restricted the usage of the scale. To sum up, scale development studies should document the steps for the deductive approach similar to PRISMA guidelines for systematic literature review [46] and provide in-depth insights for the inductive approach. For future work, we propose looking into the best practices of inductive and deductive approaches to ensure broad and holistic coverage for item generation.

Role of Background of Researchers: Researchers may also play a role in the scale development process, specifically in the item formulation stage. For example, when generating an initial pool of items for testing, the researchers may not select something that appears less important to them because of their cultural background. Still, it may seem as important in other cultures. This has been widely seen, especially in the case of privacy. An example of privacy violation is shoulder surfing–a privacy threat–which has been reported to be perceived as less concerning by high socioeconomic groups such as Germany and severely concerning by low socio-economic groups such as Egypt [48]. Therefore, there lies a risk of the researcher's background impacting the selection of the initial pool of items. In sum, the initial pool of items should be cross-validated by researchers from different cultural backgrounds. For future work, we propose investigating how the potential biases of the role of the background of researchers can be overcome in the item generation phases.

Translations of Privacy Scales: Language plays a significant role in how statements are perceived and understood by the general public. The PDBS scale was developed in an English-speaking country (Australia), and our study participants belonged to non-English-speaking countries. The language difference may have caused misinterpretation of the statements and resulted in poor model fit. While there exist translations of scales in other languages [38], the translations of privacy scales are still missing. Making translations available for privacy scales is crucial to ensure accurate measurement. Therefore, researchers should consider developing a translation of scales for use in other languages. For future work, we propose exploring what new challenges can arise for scale validity when translating scales into other languages.

Validation Assessments: Scale validation is an ongoing process which must be ensured over time and across populations to ensure accurate measurement. While many scales in literature have been developed through rigorous methodology and tested with large population samples, there still lies a need for their validation with time. For example, scales developed in the early 2000s would now need to be re-tested and validated for accurate measurement in 2025. This is so because of the evolution of different factors such as technology, which evolves and takes leaps forward, or politics, which shifts slowly and quite suddenly. Furthermore, the mode of presentation of the scale questionnaire also needs to be taken care of. For example, previously, questionnaire data used to be collected on paper; with the use of online technology, questionnaire data is being collected electronically. The mode of questionnaire data collection may impact how questionnaires are filled in, which ultimately impacts validity. In the same line of research, recent work has presented preliminary evidence that the choice of response widgets, such as radio buttons or dropdowns, impacts the scale validity and reliability [22]. This shows that many factors, from validation over time to the mode of data collection and questionnaire presentation, can impact the reliability and validity of questionnaires. Thus, along with the development and validation of questionnaire scales, the presentation of the questionnaire should also be inspected to identify and address any potential variable that may impact the results' validity and reliability. For future work, we propose inspection of what are the best design and validation guidelines for scale questionnaires.

5 Conclusion

Psychometric scales are important measuring tools in human-computer interaction. Measuring granular concepts requires scales to be accurate, reliable, and valid over time and across populations. In this study, we examined the external validity of the PDBS across 16 countries in the MENA region. Our results provide evidence of poor model fit and insufficient reliability and validity. This shows that despite careful construction of the PDBS, the PDBS lacked reliability and validity and did not achieve the required model fit indices when administered in other cultures. The results call for an immediate assessment of the scales when administered in different cultural contexts to ensure accurate and precise measurements.

Acknowledgments

This research was funded by Meta Research through the People's Expectations and Experiences With Digital Privacy Program (grant # 1014828205634642).

References

[1] Irwin Altman. 1977. Privacy regulation: Culturally universal or culturally specific? *Journal of social issues* 33, 3 (1977), 66–84.

[2] Duane F Alwin. 2007. *Margins of error: A study of reliability in survey measurement.* John Wiley & Sons.

[3] Ruwan Bandara, Mario Fernando, and Shahriar Akter. 2021. Managing consumer privacy concerns and defensive behaviours in the digital marketplace. *European Journal of Marketing* 55, 1 (2021), 219–246.

[4] David Banisar. 2011. The right to information and privacy: balancing rights and managing conflicts. *World Bank Institute Governance Working Paper* (2011).

[5] Peter M Bentler. 1990. Comparative fit indexes in structural models. *Psychological bulletin* 107, 2 (1990), 238.

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

[7] Mark J Brandt, Hans IJzerman, Ap Dijksterhuis, Frank J Farach, Jason Geller, Roger Giner-Sorolla, James A Grange, Marco Perugini, Jeffrey R Spies, and Anna Van't Veer. 2014. The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology* 50 (2014), 217–224.

[8] Richard W Brislin. 1970. Back-translation for cross-cultural research. *Journal of cross-cultural psychology* 1, 3 (1970), 185–216.

[9] Moira Burke and Robert E Kraut. 2016. The relationship between Facebook use and well-being depends on communication type and tie strength. *Journal of computer-mediated communication* 21, 4 (2016), 265–281.

[10] John T Cacioppo, Robert M Kaplan, Jon A Krosnick, James L Olds, and Heather Dean. 2015. Social, behavioral, and economic sciences perspectives on robust and reliable science. *Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences* 1 (2015).

[11] Paul Cairns. 2019. *Doing better statistics in human-computer interaction.* Cambridge University Press.

[12] Paul Antony Cairns. 2016. Experimental methods in human-computer interaction. In *Encylopedia of Human-Computer Interaction, 2nd edn.* Interaction Design Foundation, 34.

[13] Barbara A Cerny and Henry F Kaiser. 1977. A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate behavioral research* 12, 1 (1977), 43–47.

[14] Hichang Cho, Bart Knijnenburg, Alfred Kobsa, and Yao Li. 2018. Collective privacy management in social media: A cross-cultural validation. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 3 (2018), 1–33.

[15] Hichang Cho, Milagros Rivera-Sánchez, and Sun Sun Lim. 2009. A multinational study on online privacy: global concerns and local responses. *New media & society* 11, 3 (2009), 395–416.

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

[17] Christian S Crandall and Jeffrey W Sherman. 2016. On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology* 66 (2016), 93–99.

[18] Lee J Cronbach. 1950. Further evidence on response sets and test design. *Educational and psychological measurement* 10, 1 (1950), 3–31.

[19] Nicole B Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of computer-mediated communication* 12, 4 (2007), 1143–1168.

[20] Ali Farooq, Laila Dahabiyeh, and Christian Maier. 2023. Social media discontinuation: A systematic literature review on drivers and inhibitors. *Telematics and Informatics* 77 (2023), 101924.

[21] Ali Farooq, Joni Salminen, Justin Martin, Kholoud Aldous, Soon-Gyo Jung, and Bernard J Jansen. 2024. Exploring social media privacy concerns: a comprehensive survey study across 16 Middle Eastern and North African countries. *IEEE Access* (2024).

[22] Habiba Farzand, David Al Baiaty Suarez, Thomas Goodge, Shaun Alexander Macdonald, Karola Marky, Mohamed Khamis, and Paul Cairns. 2024. Beyond Aesthetics: Evaluating Response Widgets for Reliability & Construct Validity of Scale Questionnaires. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems.* 1–7.

[23] Habiba Farzand, Karola Marky, and Mohamed Khamis. 2024. Out-of-device privacy unveiled: Designing and validating the out-of-device privacy scale (odps). In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–15.

[24] Erika George. 2017. Corporate social responsibility and social media corporations: Incorporating human rights through rankings, self-regulation and shareholder resolutions. *Duke J. Comp. & Int'l L.* 28 (2017), 521.

[25] Thomas Groß. 2021. Validity and reliability of the scale internet users' information privacy concerns (iuipc). *Proceedings on Privacy Enhancing Technologies* (2021).

[26] Lisa Harris and Alan Rae. 2009. Social networks: the future of marketing for small business. *Journal of business strategy* 30, 5 (2009), 24–31.

[27] Rakibul Hasan, Rebecca Weil, Rudolf Siegel, and Katharina Krombholz. 2023. A psychometric scale to measure individuals' value of other people's privacy (VOPP). In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–14.

[28] Ayako A Hasegawa, Daisuke Inoue, and Mitsuaki Akiyama. 2024. How WEIRD is usable privacy and security research?. In *33rd USENIX Security Symposium.*

[29] Louis Henkin. 1974. Privacy and autonomy. *Columbia Law Review* 74 (1974), 1410.

[30] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and brain sciences* 33, 2-3 (2010), 61–83.

[31] Timothy R Hinkin. 1998. A brief tutorial on the development of measures for use in survey questionnaires. *Organizational research methods* 1, 1 (1998), 104–121.

[32] Haiyan Jia and Heng Xu. 2016. Autonomous and interdependent: Collaborative privacy management on social networking sites. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* 4286–4297.

[33] Mohsen Jozani, Emmanuel Ayaburi, Myung Ko, and Kim-Kwang Raymond Choo. 2020. Privacy concerns and benefits of engagement with social media-enabled apps: A privacy calculus perspective. *Computers in Human Behavior* 107 (2020), 106260.

[34] Andreas M Kaplan and Michael Haenlein. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons* 53, 1 (2010), 59–68.

[35] Hyunho Kim, Boncho Ku, Jong Yeol Kim, Young-Jae Park, Young-Bae Park, et al. 2016. Confirmatory and exploratory factor analysis for validating the phlegm pattern questionnaire for healthy subjects. *Evidence-Based Complementary and Alternative Medicine* 2016 (2016).

[36] Hanna Krasnova and Natasha F Veltri. 2010. Privacy calculus on social networking sites: Explorative evidence from Germany and USA. In *2010 43rd Hawaii international conference on system sciences.* IEEE, 1–10.

[37] Airi Lampinen, Vilma Lehtinen, Asko Lehmuskallio, and Sakari Tamminen. 2011. We're in it together: interpersonal management of disclosure in social network services. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 3217–3226.

[38] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *Proceedings of the Symposium of the Austrian HCI and Usability Engineering Group (USAB).* Springer, Cham,

Switzerland, 63–76. https://doi.org/10.1007/978-3-540-89350-9_6
[39] Yao Li, Reza Ghaiumy Anaraky, and Bart Knijnenburg. 2021. How not to measure social network privacy: A cross-country investigation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–32.
[40] May Lwin, Jochen Wirtz, and Jerome D Williams. 2007. Consumer online privacy concerns and responses: a power–responsibility equilibrium perspective. *Journal of the Academy of Marketing Science* 35 (2007), 572–585.
[41] Robert C MacCallum, Michael W Browne, and Hazuki M Sugawara. 1996. Power analysis and determination of sample size for covariance structure modeling. *Psychological methods* 1, 2 (1996), 130.
[42] Naresh K Malhotra, Sung S Kim, and James Agarwal. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research* 15, 4 (2004), 336–355.
[43] Fabiane FR Morgado, Juliana FF Meireles, Clara M Neves, Ana Amaral, and Maria EC Ferreira. 2017. Scale development: ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica* 30 (2017).
[44] Richard G Netemeyer, William O Bearden, and Subhash Sharma. 2003. *Scaling procedures: Issues and applications.* sage publications.
[45] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.
[46] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, et al. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj* 372 (2021).
[47] Clay Posey, Paul Benjamin Lowry, Tom L Roberts, and T Selwyn Ellis. 2010. Proposing the online community self-disclosure model: the case of working professionals in France and the UK who use online communities. *European journal of information systems* 19, 2 (2010), 181–195.
[48] Mennatallah Saleh, Mohamed Khamis, and Christian Sturm. 2019. What about my privacy, Habibi? Understanding privacy concerns and perceptions of users from different socioeconomic groups in the Arab World. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part III 17*. Springer, 67–87.
[49] Anna Cinzia Squicciarini, Mohamed Shehab, and Federica Paci. 2009. Collective privacy management in social networks. In *Proceedings of the 18th international conference on World wide web*. 521–530.
[50] Wolfgang Stroebe and Fritz Strack. 2014. The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science* 9, 1 (2014), 59–71.
[51] Sabine Trepte, Michael Scharkow, and Tobias Dienlin. 2020. The privacy calculus contextualized: The influence of affordances. *Computers in Human Behavior* 104 (2020), 106115.
[52] Alan F Westin. 1968. Privacy and freedom. *Washington and Lee Law Review* 25, 1 (1968), 166.
[53] Belle Wong. 2023. *Top Social Media Statistics And Trends of 2024*. https://www.forbes.com/advisor/business/social-media-statistics/

A Appendix

This section is organised as: Table A1 shows the original Privacy Defensive Behavior Scale by Bandara et al. [3]. Table A2 presents the demographic details of the participants recruited for this study. Table A3 below shows the results of Confirmatory Factor Analysis (CFA) after removing the items with the lowest loadings across all countries. To check for the suitability of the dataset for performing factor analysis, we ran the KMO MSA test. The results are as follows:

(1) Algeria (KMO=0.682, x^2= 384.767, ρ<.001),
(2) Egypt (KMO=0.703, x^2=406.681, 384.767, ρ<.001),
(3) Libya (KMO=0.693, x^2 = 239.429, ρ<.001),
(4) Morocco (KMO=0.689, x^2 = 400.250, ρ<.001),
(5) Tunisia (KMO=0.696, x^2 = 404.534, ρ<.001),
(6) Iraq (KMO=0.729, x^2 = 326.996, ρ<.001),
(7) Jordan (KMO=0.727, x^2 = 381.945, ρ<.001),
(8) Lebanon (KMO=0.690, x^2 = 298.700, ρ<.001),
(9) Palestine (KMO=0.688, x^2 = 266.582, ρ<.001),
(10) Bahrain (KMO=0.673, x^2 = 281.232, ρ<.001),
(11) Kuwait (KMO=0.717, x^2 = 372.213, ρ<.001),
(12) Oman (KMO=0.677, x^2 = 330.693, ρ<.001),
(13) Qatar (KMO=0.718, x^2 = 460.206, ρ<.001),
(14) KSA (KMO=0.733, x^2 = 497.966, ρ<.001),
(15) UAE (KMO=0.690, x^2 = 306.141, ρ<.001), and
(16) Yemen (KMO=.635, x^2 = 317.154, ρ<.001).

Item	Item Statement	Loading	AVE	CR
PDBS_1	I falsify some of my personal information when asked by social media	0.779	0.592	0.897
PDBS_2	I provide incomplete information to social media service providers	0.707		
PDBS_3	I take measures to prevent social media providers from tracking my browsing behaviour	0.772		
PDBS_4	I use software or applications to protect online privacy	0.781		
PDBS_5	I refuse to give information to online companies when I think it is too personal	0.772		
PDBS_6	I use social media platforms that do not ask for too much information	0.802		

Table A1: The Table shows the results from the original paper, including statements of the Privacy Defensive Behaviour, item loadings, Composite Reliability and Average Variance Extracted.

Country	Total (n)	Male %	Female %	Age		Local %	Expats %	University Degree %
				M	SD			
Algeria	551	52	48	37.27	12.15	95	5	77
Egypt	552	53	47	30.11	8.87	96	4	87
Libya	486	72	28	36.38	12.51	77	23	52
Morocco	530	53	47	26.84	8.24	92	8	64
Tunisia	574	55	45	33.27	11.32	90	10	76
Iraq	526	65	35	31.15	8.99	79	21	60
Jordan	580	54	46	33.41	11	87	13	79
Lebanon	485	53	47	27.69	8.31	77	23	68
Palestine	486	66	34	36.77	12.35	88	12	70
Bahrain	453	56	44	30.6	8.43	50	50	64
Kuwait	459	59	41	32.65	8.34	41	59	70
Oman	471	58	42	36.38	10.32	47	53	69
Qatar	489	69	31	35.09	11.34	30	70	71
Saudi Arabia	521	61	39	35.78	9.8	55	45	75
UAE	479	72	28	29.75	8.27	11	89	78
Yemen	528	89	11	29.06	9.21	90	10	55
Total	**8140**	**61**	**39**	**32.74**	**10.67**	**70**	**30**	**69**

Table A2: Demographic details of the Participants

	Fit Indices after eliminating PDBS_1			Fit Indices after eliminating PDBS_4			Fit Indices after Removing PDBS_5		
	CFI	**TLI**	**RMSEA**	**CFI**	**TLI**	**RMSEA**	**CFI**	**TLI**	**RMSEA**
Algeria	0.898	0.694	0.086	0.922	0.765	0.092	0.837	0.51	0.128
Egypt	0.916	0.749	0.088	0.92	0.759	0.088	0.795	0.386	0.149
Libya	0.997	0.991	0.013	0.901	0.703	0.081	0.849	0.548	0.093
Morocco	0.911	0.732	0.085	0.887	0.66	0.11	0.915	0.745	0.061
Tunisia	0.879	0.638	0.112	0.904	0.712	0.101	0.752	0.257	0.156
Iraq	0.928	0.783	0.07	0.966	0.898	0.056	0.93	0.79	0.075
Jordan	0.941	0.882	0.077	0.958	0.916	0.067	0.926	0.852	0.081
Lebanon	0.942	0.827	0.063	0.873	0.62	0.098	0.806	0.419	0.117
Palestine	0.946	0.837	0.06	0.923	0.768	0.071	0.819	0.458	0.103
Bahrain	0.986	0.959	0.026	0.837	0.51	0.112	0.884	0.651	0.095
Kuwait	0.967	0.901	0.054	0.839	0.516	0.126	0.779	0.336	0.139
Oman	0.926	0.779	0.07	0.835	0.506	0.122	0.811	0.434	0.124
Qatar	0.98	0.94	0.045	0.777	0.33	0.168	0.853	0.559	0.13
KSA	0.944	0.832	0.08	0.907	0.722	0.115	0.861	0.582	0.133
UAE	0.911	0.734	0.078	0.862	0.585	0.102	0.812	0.436	0.119
Yemen	0.792	0.376	0.121	0.826	0.478	0.118	0.722	0.167	0.145

Table A3: The Table presents the results of the Confirmatory Factor Analysis when items with low loadings are removed. The cells highlighted with a green background indicate where the threshold for the fit indices was achieved.