German Translation of the Concerns for Information Privacy (CFIP) Construct

David Harborth and Sebastian Pape

Goethe University, Chair of Mobile Business & Multilateral Security,
Theodor-W.-Adorno Platz 4, 60323 Frankfurt, Germany,
david.harborth@m-chair.de

January 29, 2018

Abstract

We present and validate a German translation of the construct *Concerns for Information Privacy (CFIP)*. This construct, consisting of four sub-constructs, measures the privacy concerns of individuals with regard to organiziational privacy practices. With this scope, the construct has a wide applicability for quantitative research on privacy. We surveyed participants on the location-based mobile augmented reality game Pokémon Go. We conducted the translation with the help of two independent and certified translators and tested the validity and reliability of the constructs by conducting a confirmatory factor analysis (CFA). The analysis is based on a sample of 681 active players of the game. The participants were acquired with the help of a certified German panel provider. The results indicate the validity and reliability of the German translation of the CFIP construct for the case of Pokémon Go. Professional translations of existing constructs are necessary to apply established models and associated questionnaires in other countries. This holds in particular, because language may influence survey responses, especially with regard to attitudes. However, these translations are associated with high monetary costs and efforts and seldom published. Therefore, we provide opportunities for future work by making our valid and reliable German translation of the CFIP construct accessible to interested researchers.

Keywords: privacy concerns, certified translation, open data, privacy measures, German translation

1 Introduction

Dealing with privacy issues is a highly relevant issue in today's digitalized world. An important part to understand users' perceptions about privacy are the concerns of people related to data practices of organizations. One of the most established questionnaire to measure these concerns quantitatively is the instrument *Concerns for Information Privacy (CFIP)* by Smith et al. [21]. The instrument is divided into four subscales or constructs (Collection, Errors, Unauthorized Secondary Use, Improper Access). Collection is defined as the concern of people that too much data about them is collected over time. Errors represent users' concerns about inaccurate or false personal data in databases. Unauthorized secondary use measures the concern that personal data is used for another purpose than initially disclosed without the user's authorization. Improper access captures concerns about unauthorized people having access to the user's personal data [21]. As these constructs cannot be measured directly (latent variables), they have to be operationalized in order to quantify the concerns for a user study. This operationalization is done via a questionnaire for quantitative surveys (cf. Table 1). The CFIP instrument is used in the literature in several different methodological ways. For example, the four constructs can be analysed individually by forming sum scores of the answers of participants to determine how concerned a person is with regard to a specific organization or company [8]. Other literature uses the four constructs to form a second-order construct *CFIP* and use this as a predictor for several analyses [22]. A deep statistical evaluation of the different methods to investigate constructs in this paper is beyond the scope of this work. For more details we refer to Malhotra et al. [14, 15], Steward and Segars [22] and Jarvis et al. [12].

Bellman et al. [1] conduct a study with users from 38 countries and find that cultural values were associated with differences in privacy concerns. But even within Western cultures Whitman [25] notes differences in sensibilities about what ought to be kept "private". For further insights how cultural factors influence privacy concerns we refer to Li et al. [13]. Directly connected to cultural differences are differences in privacy regulations, where the history about the Safe Harbor Agreement (cf. Weiss and Archick [24]) supports the view that differences within Western cultures exist. Since different privacy regulations are applied in different locations, the need arises to challenge existing results in various locations respectively cultures. This is in particular true for Germany. Singh and Hill [20] find that "...consumers in Germany have very strong views about protecting their privacy. They believe that both companies and governments are obligated to protect the information of their consumers and citizens."

However, directly connected to the location is the language spoken at a specific location. The problem of translating surveys is not new. Ervin and Bower [3] state that while in theory methodological considerations should be the only ones, in practice other matters like costs of translation cannot be ignored. However, Pérez [17] describes how language effects survey responses and Ogunnaike et al. [16] show that language may implicitly influence attitudes with a large effect size (d=.72). Most scientific articles employing the CFIP constructs do not mention the translation or do not describe the translation process and its result if the study is conducted in another language (cf. examples for Germany [5, 10]). However, Ervin and Bower [3] point out the importance of a rigorous translation process. We argue that it is not necessary to do a professional translation each time again. In particular, for an often used and well established construct, researchers should build on the results of previous studies and either spare the efforts and costs of the translation or invest it in improving existing translations and its validations. In general, the research subject - in our case Pokémon Go - can easily be substituted with the respective technology of a certain organization. Thus, by publishing the translation and validating it, we contribute a generally applicable German questionnaire of the CFIP construct and allow other researchers to build on our work. The remainder of the paper is as follows. Section 2 describes the methodology and the translation process in detail. The translated questionnaire is presented here. Reliability and validity tests can be found in Section 3. We end with a brief discussion and conclusion of the results in Section 4.

2 Methodology

In the following subsections, we discuss the questionnaire translation and the data collection process. This section also contains the German translation of the CFIP instrument in Table 1.

2.1 Questionnaire Translation

To ensure content validity of the translation, we followed a rigorous translation process proposed by Venkatesh et al. [23]. First, we translated the English questionnaire into German with the help of a translator certified by the DIN EN 15308 norm. EN 15038 was defined in 2006 by CEN, the European Committee for Standardization [4] and is a quality standard developed especially for translation services providers [26]. Although, DIN EN 15038 was superseded by ISO 17100 in November 2015, DIN EN 15038 is still in place and it will take a while until certifications are fully replaced by ISO17100 [11].

The German version was then given to a second certified translator who independently retranslated the questionnaire to English. This step was done to ensure the equivalence of the translation. Third, a group of five academic colleagues checked the two English versions with regard to this equivalence. All items were found to be equivalent, except for one. For this case, we contacted the translator of the German version and discussed and solved the issue personally. In a last step, the German version of the questionnaire was administered to students of a Master's course to check preliminary reliability and validity. The original items by Smith et al. [21] and the German translation can be found in Table 1. The single items of the questionnaire are supposed to be answered on a 7-point Likert scale ranging from strongly disagree to strongly agree.

Table 1: German Questionnaire Translation

Construct	Original English Items	German Translation
Collection	COLL1. It usually bothers me when companies ask me for personal information.	COLL1-G. Ich mache mir für gewöhnlich Gedanken darüber, wenn Unternehmen mich nach meinen persönlichen Informationen fragen.
	COLL2. When companies ask me for personal information, I sometimes think twice before providing it.	COLL2-G. Ich denke manchmal zweimal darüber nach, meine persönlichen Daten auszuhändigen, wenn Unternehmen mich danach fragen.
	COLL3. It bothers me to give personal information to so many companies.	COLL3-G. Es stört mich, meine persönlichen Informationen an so viele Unternehmen weiterzugeben.
	COLL4. I am concerned that companies are collecting too much personal information about me.	COLL4-G. Ich mache mir Sorgen, dass Unternehmen zu viele persönliche Daten von mir sammeln.
Errors	ERR1. All the personal information in computer databases should be double-checked for accuracy – no matter how much this costs.	ERR1-G. Alle persönlichen Informationen in Computerdatenbanken sollten ungeachtet der dafür entstehenden Kosten doppelt auf Genauigkeit hin überprüft werden.
	ERR2. Companies should take more steps to make sure that the personal information in their files is accurate.	ERR2-G. Unternehmen sollten mehr Schritte unternehmen, sicherzustellen, dass die persönlichen Informationen in ihren Dateien korrekt sind.
	ERR3. Companies should have better procedures to correct errors in personal information.	ERR3-G. Unternehmen sollten bessere Verfahren anwenden, um Fehler bei persönlichen Informationen zu korrigieren.
	ERR4. Companies should devote more time and effort to verifying the accuracy of the personal information in their databases.	ERR4-G. Unternehmen sollten mehr Zeit und Bemühungen investieren, um die Genauigkeit persönlicher Informationen in ihren Datenbanken zu überprüfen.

Table 1: German Questionnaire Translation

Construct	Original English Items	German Translation
Unauthorized Secondary Use	USU1. Companies should not use personal information for any purposes unless it has been authorized by the individuals who provided the information.	USU1-G. Unternehmen sollten persönliche Informationen nicht zu anderen Zwecken verwenden, sofern diesem nicht seitens der Einzelpersonen, die die Informationen bereitgestellt haben, zugestimmt wurde.
	USU2. When people give personal information to a company for some reason, the company should never use the information for any other reason.	USU2-G. Wenn Personen aus irgendeinem Grund persönliche Informationen an ein Unternehmen weitergeben, sollte das Unternehmen diese Informationen niemals zu anderen Zwecken verwenden.
	USU3. Companies should never sell the personal information in their computer databases to other companies.	USU3-G. Unternehmen sollten die in ihren Computerdatenbanken hinterlegten persönlichen Daten niemals an andere Unternehmen verkaufen.
	USU4. Companies should never share personal information with other companies unless it has been authorized by the individuals who provided the information.	USU4-G. Unternehmen sollten persönliche Informationen niemals an andere Unternehmen weitergeben, sofern diesem nicht seitens der Einzelpersonen, die die Informationen bereitgestellt haben, zugestimmt wurde.
Improper Access	IA1. Companies should devote more time and effort to preventing unauthorized access to personal information.	IA1-G. Unternehmen sollten mehr Zeit und Bemühungen investieren, um unautorisiertem Zugriff auf persönliche Daten vorzubeugen.
	IA2. Computer databases that contain personal information should be protected from unauthorized access – no matter how much it costs.	IA2-G. Computerdatenbanken, die persönliche Informationen enthalten, sollten ungeachtet der dafür entstehenden Kosten vor unautorisiertem Zugriff geschützt werden.
	IA3. Companies should take more steps to make sure that unauthorized people cannot access personal information in their computers.	IA3-G. Unternehmen sollten mehr Schritte unternehmen, um sicherzustellen, dass unautorisierte Personen nicht auf persönliche Informationen auf ihren Computern zugreifen können.

2.2 Data Collection and Demographics

We decided to conduct the study with the help of a German sample provider to have representative sample. Thereby, we could ensure two things. To ensure quality of our data, we chose a certified provider (certified following the ISO 26362 norm). We installed the survey on a university server and managed it with the survey software LimeSurvey (version 2.63.1) [19]. This link was distributed by the panel provider to 9338 participants. Of those 9338 approached participants, only 681 remained after asking whether they play Pokémon Go, whether they are older than 18 years old and, whether they answered a test question in the middle of the survey correctly. Besides this test question, we asked the Pokémon Go players about their current level. We designed this question intentionally as a free field question with numeric entries only. As Pokémon Go ends at level 40. we could test the knowledge of the participants and establish an additional screen-out mechanism. We sorted out all participants who stated to have a level higher than 40. Since they were actually not playing, they did not answer the questions carefully or they did not take the questionnaire seriously enough. In addition, two participants stated that they "never" play Pokémon Go.

3 Reliability and Validity Tests

To test the validity and reliability of our translation, we need to analyse the instrument with a confirmatory factor analysis (CFA). This type of analysis is usually done if the instrument and constructs under investigation are known [2]. For conducting the CFA, we use SmartPLS version 3.2.6 [18]. For the PLS algorithm, we choose the factor weighting scheme with a maximum of 300 iterations and a stop criterion of 10^{-7}. For the bootstrapping procedure, we use 5000 bootstrap subsamples and no sign changes as the method for handling sign changes during the iterations of the bootstrapping procedure.

To assess the German translation of the questionnaire, we conduct the following statistical analyses. First, we analyse the internal consistency reliability (ICR). After that, we assess the convergent and discriminant validity. All these tests belong to the necessary steps of evaluating the measurement model with reflective constructs [7]. The last analysis belongs to the structural model assessment, whereas it is also very important for the translation itself. We assess whether there are substantial correlations among the constructs themselves (collinearity). If this was the case for our translated version, the model would not be measuring the results correctly. Therefore, we also test for collinearity.

3.1 Internal Consistency Reliability

The internal consistency reliability (ICR) indicates how well certain indicators of a construct measure the same latent phenomenon. Two standard approaches for assessing ICR are Cronbach's α and the composite reliability. The values of both measures should be between 0.7 and 0.95 for research that builds upon accepted models. Values of Cronbach's α are seen as a lower bound and values of the composite reliability as an upper bound of the assessment [7]. Table 2 includes the ICR of the used variables in the last two rows. It can be seen that all values for both measures are above the lower threshold of 0.7 and below the upper threshold of 0.95. Values above that upper threshold indicate that the indicators measure the same dimension of the latent variable, which is not optimal with regard to the validity [7]. Based on these results ICR is established for the translated version of the CFIP instrument.

Table 2: Loadings and cross-loadings of the reflective items and ICR measures

Items	Collection	Errors	Improper Access	Unauthorized Secondary Use
COLL1	**0.833**	0.397	0.508	0.514
COLL2	**0.831**	0.413	0.574	0.528
COLL3	**0.832**	0.371	0.483	0.469
COLL4	**0.815**	0.405	0.427	0.416
ERR1	0.429	**0.840**	0.440	0.421
ERR2	0.409	**0.887**	0.481	0.436
ERR3	0.432	**0.860**	0.521	0.460
ERR4	0.373	**0.852**	0.420	0.374
IPA1	0.569	0.508	**0.915**	0.791
IPA2	0.516	0.493	**0.886**	0.753
IPA3	0.551	0.469	**0.900**	0.763
USU1	0.553	0.461	0.796	**0.891**
USU2	0.518	0.450	0.748	**0.880**
USU3	0.510	0.399	0.739	**0.872**
USU4	0.472	0.424	0.718	**0.873**
Cronbach's α	0.847	0.883	0.884	0.902
Comp. Reliability	0.897	0.919	0.928	0.932

3.2 Convergent Validity

Convergent validity determines the degree to which indicators of a certain reflective construct are explained by that construct. This is assessed by calculating the outer loadings of the indicators of the constructs (indicator reliability) and by looking at the average variance extracted (AVE) [6]. Loadings above 0.7 imply that the indicators have much in common, which is desirable for reflective measurement models [7]. Table 2 shows the outer loadings in bold on the diagonal. All loadings are higher than 0.7. The AVE indicates convergent validity for a construct as a whole. AVE is equal to the sum of the squared loadings divided by the number of indicators. A threshold of 0.5 is acceptable, indicating that the construct explains at least half of the variance of the indicators [7]. The first column of Table 3 presents the AVE of the constructs in parentheses. All values are above 0.5, demonstrating convergent validity.

Table 3: Convergent (AVEs) and discriminant validity (Fornell-Larcker approach

Sub-Constructs (AVE)	Collection	Errors	Improper Access	Unauthorized Secondary Use
Collection (0.685)	0.828			
Errors (0.739)	0.479	0.860		
Improper Access (0.811)	0.606	0.544	0.901	
Unauthorized Secondary Use (0.773)	0.585	0.494	0.854	0.879

3.3 Discriminant Validity

Discriminant validity measures the degree of uniqueness of a construct compared to other constructs. Comparable to the convergent validity assessment, two approaches are used for investigated discriminant validity. The first approach, assessing cross-loadings, is dealing with single indicators. All outer loadings of a certain construct should be larger than its cross-loadings with other constructs [6]. Table 2 illustrates the cross-loadings as off-diagonal elements. All cross-loadings are smaller than the outer loadings, fulfilling the first assessment approach of discriminant validity. The second approach is on the construct level and compares the square root of the constructs' AVE with the correlations with other constructs. The square root of the AVE of a single construct should be larger than the correlation with other constructs (Fornell-Larcker criterion) [7]. Table 3 contains the square root of the AVE on the diagonal in parentheses. All values are larger than the correlations with other constructs, indicating discriminant validity. Since there are problems in determining the discriminant validity with both approaches, researchers propose the heterotrait-monotrait ratio (HTMT) for assessing discriminant validity as a superior approach to the others [9]. HTMT divides between-trait correlations by within-trait correlations, therefore providing a measure of what the true correlation of two constructs would be if the measurement is flawless. Values close to 1 for HTMT indicate a lack of discriminant validity. A conservative threshold is 0.85. Table 4 contains the values for HTMT. No value is above the suggested threshold of 0.85 except for the HTMT ratio between improper access and unauthorized secondary use. The ratio for these constructs is 0.956, indicating a lack of discriminant validity. For a further analysis to test whether the translation lacks validity, a follow-up analysis related to HTMT must be conducted.

To evaluate whether the HTMT statistics are significantly different from 1, a bootstrapping procedure with 5,000 subsamples is conducted to get the confidence interval in which the true HTMT value lies with a 95% chance. The HTMT measure requires that no confidence interval

Table 4: Discriminant validity (HTMT approach)

Constructs	Collection	Errors	Improper Access
Errors	0.552		
Improper Access	0.694	0.613	
Unauthorized Secondary Use	0.664	0.550	0.956

Table 5: Confidence intervals for HTMT

	Original Sample (O)	Sample Mean (M)	Bias	2.5%	97.5%
ERR ->COLL	0.552	0.553	0.002	0.448	0.645
IPA ->COLL	0.694	0.694	0.000	0.629	0.756
IPA ->ERR	0.613	0.614	0.001	0.528	0.694
UNSU ->COLL	0.664	0.664	0.000	0.595	0.727
UNSU ->ERR	0.550	0.551	0.001	0.465	0.630
UNSU ->IPA	0.956	0.956	0.000	0.921	0.984

contains the value 1, which is fulfilled (Table 5). Thus, although previous results are above the conservative threshold of 0.85, discriminant validity is established for our model.

3.4 Collinearity

Collinearity is present if two predictor variables are highly correlated with each other. To address this issue, we assess the inner variance inflation factor (inner VIF). All VIF values above 5 indicate that collinearity between constructs is present. For our model, the highest VIF is 1.520. Thus, collinearity is not an issue.

4 Discussion and Conclusion

We presented a German translation of the established privacy concerns measure CFIP (Concerns for Information Privacy) by Smith et al. [21]. We showed that internal consistency reliability, convergent validity and discriminant validity is given. The results indicate that there is no bias for our case. Furthermore, collinearity among the sub-constructs seems to be not existent. In summary, our analyses indicate that our translated version of the CFIP questionnaire is a valid and reliable instrument for future work on privacy with German speaking participants. By providing the instrument to the research community, we hope to foster research in other languages and to encourage more researchers to publish their research materials, like translations of questionnaires or raw data.

5 Acknowledgments

The authors wish to thank the Faculty of Economics and Business Administration of the Goethe University Frankfurt am Main for supporting this work with a grant within the funding program "Forschungstopf". This research was also partly funded by the German Federal Ministry of Education and Research (BMBF) with grant number: 16KIS0371.

References

[1] Steven Bellman, Eric J. Johnson, Stephen J. Kobrin, and Gerald L. Lohse. International differences in information privacy concerns: A global survey of consumers. *The Information Society*, 20(5):313–324, 2004.

[2] Timothy A. Brown. *Confirmatory Factor Analysis for Applied Research*. Guilford Press, New York, 2015.

[3] Susan Ervin and Robert T. Bower. Translation Problems in International Surveys. *In Public Opinion Quaterly*, 16(4):595–604, 1952.

[4] European Committee for Standardization. CEN 2017.

[5] Mark Fodor and Alexander Brem. Do privacy concerns matter for millennials? results from an empirical analysis of location-based services adoption in germany. *Computers in Human Behavior*, 53(Supplement C):344 – 353, 2015.

[6] Joe F. Hair, Christian M. Ringle, and Marko Sarstedt. PLS-SEM: Indeed a Silver Bullet. *The Journal of Marketing Theory and Practice*, 19(2):139–152, 2011.

[7] Joseph F. Hair, G. Tomas M. Hult, Christian M. Ringle, and Marko Sarstedt. SAGE Publications.

[8] David Harborth and Sebastian Pape. Privacy Concerns and Behavior of Pokémon Go Players in Germany. In *Proceedings of IFIP Summer School on Privacy and Identity Management (IFIPSC2017)*, pages 1–18, 2017. (to appear).

[9] Jörg Henseler, Christian M. Ringle, and Marko Sarstedt. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1):115–135, 2015.

[10] Oliver Hinz, Eva Gerstmeier, Omid Tafreschi, Matthias Enzmann, and Markus Schneider. Customer loyalty programs and privacy concerns. In *20th Bled eConference: eMergence: Merging and Emerging Technologies, Processes, and Institutions, Bled, Slovenia, June 3-6, 2007*, page 32, 2007.

[11] International Organization for Standardization. Translation Services - Requirements for translation services ISO 17100:2015, 2015.

[12] Cheryl Burke Jarvis, Scott B Mackenzie, and Philip M Podsakoff. A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research*, 30(September 2003):199–218, 2003.

[13] Yao Li, Alfred Kobsa, Bart P Knijnenburg, Carolyn Nguyen, et al. Cross-cultural privacy prediction. *Proceedings on Privacy Enhancing Technologies*, 2017(2):113–132, 2017.

[14] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*, 15(4):336–355, dec 2004.

[15] Naresh K. Malhotra, Sung S. Kim, and Ashutosh Patil. Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research. *Management Science*, 52(12):1865–1883, 2006.

[16] Oludamini Ogunnaike, Yarrow Dunham, and Mahzarin R. Banaji. The language of implicit preferences. *Journal of Experimental Social Psychology*, 46(6):999–1003, nov 2010.

[17] Efrén O. Pérez. The Language-Opinion Connection. In Lonna Rae Atkeson and R. Michael Alvarez, editors, *The Oxford Handbook of Polling and Survey Methods*. Oxford uni edition, 2015.

[18] Christian M. Ringle, S. Wende, and Jan Michael Becker. SmartPLS 3, 2015.

[19] Carsten Schmitz. LimeSurvey Project Team, 2015.

[20] Tanuja Singh and Mark E Hill. Consumer privacy and the Internet in Europe: a view from Germany. *Journal of consumer marketing*, 20(7):634–651, 2003.

[21] H. Jeff Smith, Sandra J. Milberg, and Sandra J. Burke. Information privacy: measuring individuals concerns about organizational practices. *MIS quaterly*, 20(2):167–196, 1996.

[22] Kathy A. Stewart and Albert H. Segars. An Empirical Examination of the Concern for Information Privacy Instrument. *Information Systems Research*, 13(1):36–49, 2002.

[23] Viswanath Venkatesh, James Thong, and Xin Xu. Consumer Acceptance and User of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, 36(1):157–178, 2012.

[24] Martin A Weiss and Kristin Archick. *US-EU data privacy: from safe harbor to privacy shield*. Congressional Research Service, 2016.

[25] James Q Whitman. The two western cultures of privacy: Dignity versus liberty. *Yale LJ*, 113:1151, 2003.

[26] Wikipedia. EN 15038, version from October 13th 2017, 2017.

All websites have been last accessed on December. 4th, 2017.