

# Registration

## *Revisiting the Factor Structure of IUIPC-10\**

*Thomas Groß*  
Newcastle University  
United Kingdom

### General Purpose of Pre-Registrations

Pre-registrations are research statements of intention established before a sample is evaluated and statistical inferences are undertaken. A pre-registration asserts the aim of a study, including its research questions and statistical hypotheses, methods, incl. operationalization of independent variables (IVs) and dependent variables (DVs), sample and analysis specification.

The primary reason for a pre-registration lies in the fact that a statistical inference (Null Hypothesis Significance Testing) is only valid if the statistical hypotheses are fixed before the inference is undertaken. This is grounded in a  $p$ -value being a conditional likelihood contingent on the fixed null hypothesis assumed to be true. Furthermore, pre-registrations serve as a ward against questionable research practices, such as outcome-switching, hypothesizing after the results are known (HARKing), or  $p$ -hacking... it is meant to counteract the many temptations of researcher degrees of freedom.

Pre-registrations are typically committed confidentially under embargo, with an immutable timestamp. Once the corresponding study is published, the embargo is lifted.

This is an experiment registration form for the Open Science Framework (OSF)<sup>1</sup>. It is modelled according to the format of AsPredicted<sup>2</sup>.

### Context of this Pre-Registration

#### Meta-Data of Pre-Registration.

- Open Science Framework Repository: <https://osf.io/5pywm/>
- Registered Registration File: <https://osf.io/kg3f2/>—Revisit\_IUIPC-10.pdf
- Timestamp: 2020-04-28 12:54 PM
- Archived Immutable Pre-Registration: <https://osf.io/a4pw2>
- Timestamp: 2020-04-28 3:27 PM

**Peer-Reviewed Publication.** The definitive version of the study is published as:

Thomas Groß. Validity and Reliability of the Scale Internet Users' Information Privacy Concern (IUIPC). In proceedings of the 21<sup>st</sup> Privacy Enhancing Technologies Symposium (PoPETS). 2021(2), Sciendo, De Gruyter, 2021.

**ArXiv Report.** Thomas Groß. Validity and Reliability of the Scale Internet Users' Information Privacy Concern (IUIPC) [Extended Version]. arXiv:2011.11749, 2020. <https://arxiv.org/abs/2011.11749>

### 1 Structured Abstract

**Background.** IUIPC and the short form IUIPC-10 [1] have been frequently used in studies to measure user's privacy concern. Even though IUIPC has

\*Open Science Framework: <https://osf.io/5pywm/>

<sup>1</sup><https://osf.io>

<sup>2</sup><https://aspredicted.org>

already been a refinement of prior scales, it is missing a confirmative analysis of its factor structure.

**Aim.** We evaluate the IUIPC-10 factor structure, aiming at either confirming its structure or disproving it while proposing refinements.

**Method.** We use three independent samples to evaluate the factor structure of IUIPC-10 in a covariance-based confirmatory factor analysis. The first two samples A and B with a sample size of  $N_A = 226$  and  $N_B = 402$  are used to evaluate the factor structure and diagnose problems, where Sample B is designated as the main sample for CFA. A third Sample V with  $N_V = 433$  is retained as validation sample to test the properties of the confirmed model and proposed improvements. We will test for reliability, predominately internal consistency, global and local fit, considering the residuals, as well as convergent and discriminant validity.

**Anticipated Results.** We anticipate a systematic evaluation of the original IUIPC-10 questionnaire with a wide range of metrics as well as a cross-validation on an independent sample.

**Anticipated Conclusions.** We anticipate to find either a confirmation for a questionnaire frequently used in the field, or offer improvements over the existing factor structure.

## 2 State of Data Collection

*Have any data been collected for this study yet?*

- (a)  **NO** data have been collected.
- (b)  Some data have been collected, but not analyzed.
- (c)  Some data have been collected and analyzed.

*If (b) or (c), please explain briefly:*

The samples have been gathered primarily for other studies and been subjected to analyses therein. While samples A and B will be used in factor analyses to diagnose problems with IUIPC-10, Sample V will not be analysed till the validation phase on the final proposed model.

## 3 Aims

*Hypothesis: What's the main question being asked or hypothesis being tested?*

First, we test the hypothesis whether IUIPC-10 is a sound fit, which is expanded upon in the classical exact-fit, close-fit, not-close-fit, and poor-fit hypotheses. This is complemented with an inspection of fit indices and residuals.

Should the IUIPC-10 model be rejected, we diagnose apparent problems, design a proposed solution, and then evaluate that solution on an independent validation sample.

## 4 Methods

*Give a brief overview of the methods used.*

We first attempt a CFAs on IUIPC-10 on sample A and B in an attempt to confirm its factor structure. We will use global and local fit methods to evaluate that model.

Should the model be rejected, we allow for complementing the CFA diagnoses with EFA methods on sample A and B to offer ways forward in unrestricted model. Though somewhat redundant with the residuals analysis, we imagine that the EFA view on the models is more familiar to the community.

Based on these diagnostic results, we propose a final CFA model, which is first tested on samples A and B. Should these analyses indicate a good fit and satisfying solution, we will then attempt a confirmatory factor analysis of this proposed final model on the independent cross-validation Sample V.

## 5 Measurement Variables (MVs)

We use the standard items for IUIPC-10 [1], with items designated as ctrl1–3, awa1–3, and coll1–4.

## 6 Latent Variables (LVs)

As first-order latent variables, we consider corresponding Control, Awareness, and Collection. We

note that we intend to construct a second-order factor model with IUIPC as second-order latent variable.

## 7 Mediator Variables

N/A

## 8 Moderator Variables

N/A

## 9 Data Preparation

*Describe what measures will be taken to check assumptions and label outliers.*

We will check the covariance matrices of the data inputted to the CFA for positive definitiveness. We will do so with the R package matrixcalc.

We evaluate the indicator variables of the samples for extreme collinearity based on the corresponding correlation matrixes, that is, examining them for correlations of  $r \geq .90$ .

We will use the robust outlier labelling rule to detect uni-variate outliers, intending to cap outliers at the 5<sup>th</sup> or 95<sup>th</sup> percentile, not to remove cases altogether if the observations seem to come from the same population and are not data entry errors. We intend to use the Mahalanobis distance for multi-variate outliers. For computing the outlier evaluation, we prefer the R packages car and stats.

Finally, for the evaluation of multi-variate extreme non-normality, we intend to use uni-variate analyses of skew and kurtosis, drawn from the R package MVN, largely using histograms for a visual check. In case of non-normality, we would opt for robust MLM or other estimation methods considered highly robust. We do not intend to transform the samples.

## 10 Main Analyses

*Describe what analyses (e.g., t-test, repeated-measures ANOVA) you will use to test your main hypotheses.*

The main analysis will be covariance-based confirmatory factor analysis using the R package lavaan.

## 11 Secondary Analyses

*Describe what secondary analyses you plan to conduct (e.g., order or gender effects).*

(Parallel) exploratory factor analysis will be used in case the overall factor number is called in question and as an explanatory tool to generate factor loadings in a free model. We will use the R package psych for the EFAs.

## 12 Validation

*Describe what diagnostics or validation methods you plan to employ to check the soundness of the analyses.*

We intend to validate CFAs on

1. Reliability measures, such as internal consistency via Cronbach's  $\alpha$  etc. for the constructs and their items,
2. Identification and convergence,
3. Global fit, considering exact fit with
  - $\chi^2$ , even though we do not strictly expect it to be non-significant at the intended sample sizes  $N > 400$ ,
  - CFI,
  - RMSEA  $\hat{\epsilon}$ , and its 90% Confidence Interval, and related close-fit tests,
  - SRMR,
4. Local fit, especially based on inspecting the correlation and standardized covariance residuals,
5. Convergent and discriminative validity.
6. Composite reliability and AVE, for the reliability of the factor measurement of the fitted CFA.

When it comes to the comparison of two or more models we prefer the  $\chi^2$  difference test for nested models and AIC for non-nested model or models with changes in variables involved.

## 13 Sample

*Where and from whom will data be collected? How will you decide when to stop collecting data (e.g., target sample size based on power analysis or accuracy in parameter estimation, set amount of time)? If you plan to look at the data using sequential analysis, describe that here.*

The samples were collected through Prolific Academic for other studies. Sample A with  $N_A = 226$  was sampled without restrictions on participant demographics. Sample B ( $N_B = 402$ ) and Sample V ( $N_V = 433$ ) were constrained to be representative of the UK population by age and gender.

All three samples are independent from each other; for B and V this was enforced with an explicit exclusion criterion, that is, participants who entered survey B were excluded from survey V by Prolific ID, and vice versa.

We computed an *a priori* power analysis on RM-SEA tests, aiming at a level of .05 using the R package *semPower*. We note that the full IUIPC-10 model has  $n_{\text{par}} = 23$  free parameters and  $df = 32$  degrees of freedom. To reach 80% power, we would need a target sample size of  $N_{1-\beta=.80} = 317$ . The critical  $\chi^2$  is at 46.19.

## 14 Exclusion Criteria

*Who will be excluded (e.g., outliers, participant who fail manipulation check, demographic exclusions)? Will they be replaced by other participants?*

Incomplete observations and observations from participants who failed more than one attention check will be discarded without replacement. We will report their number.

## 15 Exception Handling

*Should exceptions from the planned study occur (e.g., unexpected effects observed), how will they be handled?*

Unexpected results will be declared as exploratory.

## 16 Sign-Off

Pre-registration written by (initials): T.G.

Pre-registration reviewed by (initials): T.G.

## Change Management

**2020-11-26:** The pre-registration was amended with author disclosure, publications and project acknowledgment.

## Acknowledgment

The work was supported by ERC Starting Grant CASCAdE (GA n°716980).

## References

- [1] N. K. Malhotra, S. S. Kim, and J. Agarwal. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research*, 15(4):336–355, 2004.

## A Power Analysis

```
$type
[1] "a-priori"
$alpha
[1] 0.05
$desiredBeta
[1] 0.2
$desiredPower
[1] 0.8
$impliedBeta
[1] 0.197979
$impliedPower
[1] 0.802021
$impliedAbratio
[1] 0.2525521
$impliedNCP
[1] 25.28
$fmin
[1] 0.08
$effect
```

```
[1] 0.05
$effect.measure
[1] "RMSEA"
$requiredN
[1] 317
$df
[1] 32
$chiCrit
[1] 46.19426
$rmsea
[1] 0.05
$mc
[1] 0.9607894
$bPrecisionWarning
[1] FALSE
attr(,"class")
[1] "semPower.aPriori"
```