Development of a German Instrument for Self-Perceived Data Literacy - An Algorithm-based Approach to Scale Development

Authors: Leonie Hagitte, Timo Lorenz, Manuel Völkle, Martin Schultze
Time of Preregistration: Before data collection.

Introduction

The increasing importance of competent and critical data handling in society (Carmi et al., 2020) emphasises the necessity to assess and understand one's own data literacy. Previous approaches to data literacy primarily target specific groups and professions, with few including the general population (Cui et al., 2023). Given the diversity of theoretical approaches (Cui et al., 2023), we feel there is an urgent need for a unified definition of data literacy to ensure comparability and holistic assessment.

Objective

Our aim is to derive a comprehensive definition of data literacy based on existing approaches and to develop a questionnaire for self-perceived data literacy, measuring the three core factors of the construct. While drafts for two additional factors are included as preliminary assessments for future studies, they are not the focus of this study. To achieve this, we integrate essential factors from various disciplines via a literature review to develop a theory-driven conceptualisation of constructs. Furthermore, we apply an iterative process for item creation and refinement, followed by the selection of the best items for a final scale.
We arrived at five core facets that are also very prominent in the most definitions in the literature (Cui et al., 2023). We further divided them into "consumer" facets (Comprehension,Evaluation & Integration), that are relevant for nearly every person

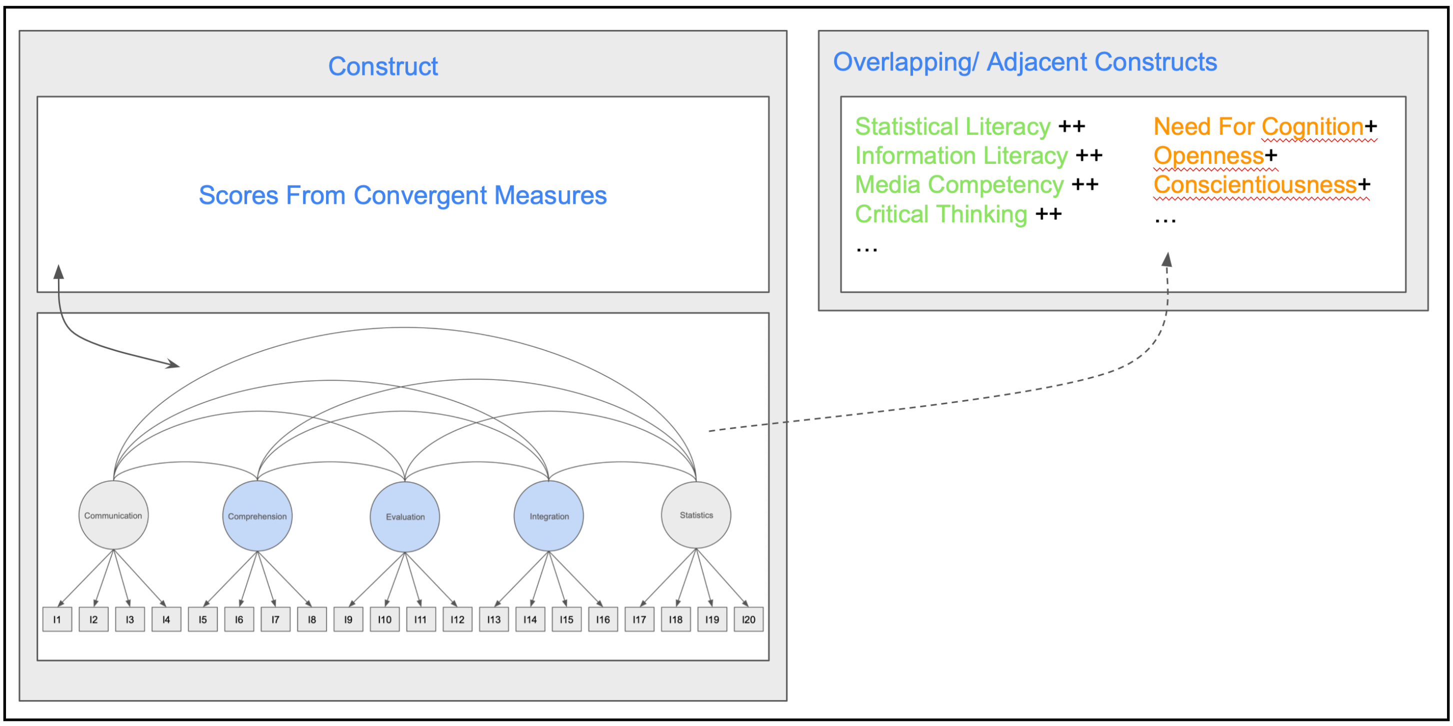

Fig.1: Nomological Network

in society and “producer” facets (Communication & Statistics), that are mainly relevant for people, actively working with data.

Participants

Sample size and source:

The participants are recruited on several online social media platforms. The participation is voluntary. We conducted a-priori power analysis to determine the necessary sample size for the structural equation modelling. We used the ‘semPower’ package in R (Moshagen & Bader, 2023) and also took a look into studies with similar goals and methods. The power analysis gave an analytical estimate for N=645, and a simulated estimate N=613, for the respective measurement model. In the literature sample sizes of N=500 up to N=1000 could be found (Algner & Lorenz, 2022; Remmert et al.,2022; Schneider et al.,2024). So the optimal sample size, we are aiming at, lies somewhere between those numbers.

Exclusion and inclusion criteria

Attention Items: Attention check questions are included (Three Instructed response items and One seriousness check item, at the end) within the survey to assess participants' attentiveness. Participants who fail to correctly answer two out of the four attention check questions will be excluded from the analysis.

Inclusion Criteria: Participants have to be of legal age, to be included in the study.

Methodology

Study Design

A cross-sectional online survey is used to examine a representative sample from the general population. Participants complete the Self-perceived Data Literacy Scale along with demographic questions and additional validation measures.

1. Item Creation:

A mixed-methods approach is employed, including a literature review to create items and cognitive interviews to refine potential items. We will treat the first 25 participants like a pilot, to check for potential problems in the survey.

2. Item Selection:

A quantitative survey is conducted, including the original item pool, demographics, and validation measures. Algorithm-based item selection is used to choose the most relevant items, reducing the item pool. The sample is split into training and test datasets to evaluate solutions against an objective function consisting of model fit criteria and composite reliability. In automated item selection, items can be chosen as sets that meet specific criteria. Those criteria will be defined in the objective function in ‘stuart’ (Schultze, 2020). We want to optimise the model to be found, for model fit and reliability (RMSEA, SRMR, CFI & McDonalds ω) as well as variability in the difficulty of items.

3. Validation:

Relationships between the newly created measure and related constructs are hypothesised. Construct validity is evaluated through confirmatory factor analysis (CFA) and correlation analyses with related constructs. Cross-validation and measurement invariance tests are also conducted.

We expect the final questionnaire to correlate moderately up to highly positive with the SWE-IV-16 (Behm, 2018), measuring peoples ability to engage in information behaviour. We expect a moderate, positive correlation of the final scale with the five general items of the ICT-SC25 (Schauffel et al., 2021). We expect a small positive correlation with the NFC-K (Beißert et al., 2015), assessing the Need for Cognition (NFC). We also expect small positive correlations with openness and conscientiousness of the BFI-10 (Rammstedt et al., 2014).

The provided sample will be divided into k subsets using the ‘holdout' function in ‘stuart’. This applies the specified item-selection procedure to the training dataset. Validation will be conducted using the ‘crossvalidate’ function in ‘stuart’, to assess the invariance of the measurement models between the training and validation datasets. Invariance levels will be measured with the ‘max.invariance' function. Invariance will be necessary to claim that the scale validation has worked.

The final item selection will be determined by the highest value on the objective function in the multiple-group SEM, while ensuring 'max.invariance' between the training and validation data.

Limitations:

The heterogeneous nature of the construct complicates global instrument development and understanding across all participants. The measure is designed for citizens, potentially limiting discrimination at higher item difficulties or among more literate participants, a direction we aim to improve in future studies.

Randomisation

Sample Allocation:

Random assignment to samples for cross-validation ensures even distribution of participants across conditions, increasing internal validity.

Survey Question Order:

Survey questions of each measurement are randomised for each participant to minimise order effects and response biases. Also the questions in each factor of the data literacy questionnaire are randomly selected for each participant. That way each participant only answers half of the possible items, the other half are planned missings. This is to shorten the overall length of the assessment and thereby to enhance the data quality. The planned, random missings will be estimated via full information maximum likelihood.

Total number of measurements:
The survey includes questions on Data Literacy, the SWE-IV-16 scale, the NFC-K scale, general items of the ICT Self-Concept Scale (ICT-SC25), openness and conscientiousness items from the BFI-10 scale, attention-checking items, and demographic questions.

1. Measure details: On Data Literacy the participants were asked to answer 71 items. Each participant will answer 38 items of the 71, that are randomly selected. To answer the items, respondents indicate their agreement on a five-point Likert scale (1 = "strongly disagree", 2 = "somewhat disagree", 3 = "neither agree nor disagree", 4 = "somewhat agree", 5 = "strongly agree”) with a “don’t know” option.

2. Measure details: The SWE-IV-16 (Behm, 2018) assesses the self-efficacy beliefs of adolescents and adults in their ability to engage in information behaviour. This questionnaire measures the process model of information-related problem-solving (Brand-Gruwel et al., 2009). It consists of 16 statements addressing self-assessed abilities in searching for and evaluating information, as well as managing information searches effectively. Each statement begins with "When I search for information on a topic or a specific question..." and respondents indicate their agreement on a five-point Likert scale (1 = "strongly disagree", 2 = "somewhat disagree", 3 = "neither agree nor disagree", 4 = "somewhat agree", 5 = "strongly agree”).
Index and scoring criteria: The total scale value is computed as the arithmetic mean of the items, which may be inverted if necessary. Calculation of the total value requires valid responses to at least 12 of the 16 items.

3. Measure details: The NFC-K (Beißert et al., 2015) is a tool used to assess the NFC through four items, which represent two facets: "engagement" and “joy". The NFC-K is measured with a seven-point response scale, ranging from "strongly disagree" (1) to "strongly agree" (7), with a "neither" option in the middle. The German version of the scale is adapted from the original English scale by Cacioppo and Petty (1982) and translated by Bless et al. (1994).
Index and scoring criteria: To determine an individual's NFC score, a mean value (scale value) is computed from the four raw score points of the responses. The resulting mean values range between 1 and 7.

4. Measure details: To assess self-perceived competence in using information and communication technology (ICT), the five general items of the ICT-SC25 (Schauffel et al., 2021) were used. The ICT-SC25 is a scale consisting of 25 items designed to assess self-perceived competence in using information and communication technology. It is available in both German (ICT-SC25g) and English (ICT-SC25e). The scale measures general and domain-specific ICT competence, including

communication, processing and storing, content generation, safe application, and problem-solving skills. Items are measured using a six-point fully-labeled Likert-type rating scale ranging from strongly disagree (1) to strongly agree (6).
Index and scoring criteria: Researchers can choose to utilise either the entire scale or individual subscales based on their specific research objectives. The ICT-SC25g/e is applicable for both manifest and latent analysis. Manifest scale scores for the ICT-SC25g/e are calculated separately for each subscale by computing the unweighted mean score of the items within each subscale (Schauffel et al., 2021).

5. Measure details: The BFI-10 (Rammstedt et al., 2014) was used to assess personality based on the five-factor model. Only the items on openness and conscientiousness were assessed.
The items were answered on a five-point rating scale from "strongly disagree" (1) to "strongly agree" (5).
Index and scoring criteria: To measure the respondent's individual traits on the five personality dimensions, the responses to the two items for each dimension are averaged. First, the negatively worded item is recoded (items 1, 3, 4, 5, and 7), then the mean value is calculated for each dimension from both the recoded and non-recoded items. The values for the five dimensions range from 1 to 5 (see Rammstedt, 2007 for reference values).

Analysis Plan

1. Research Question:
Does the proposed set of items effectively capture the latent factor structure of self-perceived data literacy, and can the created scale be considered a reliable and valid measure of this construct?
1. Hypothesis: We hypothesise that the proposed set of items will demonstrate a good model fit.
Planned Analysis: Algorithm based item selection via the R package 'stuart' (Schultze, 2020) and CFA via the R package 'lavaan' (Rosseel,2012).
Inference criteria: Model fit was assessed using established criteria (e.g.: Hu & Bentler, 1999). Comprising of $\chi 2$ significance testing as well as a combination of several fit indices, i.e.,RMSEA < 0.05, SRMR < 0.07, CFI > 0.95. Model-specific cutoff values will be considered as well, using the 'ezCutoffs' package(Schmalbach et al.,2019).
2. Hypothesis: We expect that the structure found in the initial analysis will be found as well in a different sample.
Planned Analysis: Validate the first found structure with k-fold cross-validation ('kfold' function in 'stuart') using the "crossvalidate" function of the R package"stuart" (Schultze, 2020). The crossvaligdation will follow a holdout-sample approach.

Assumptions: For item selection we used random sample split to arrive at a training-sample for the algorithm and a test-sample. The k-fold cross validations were conducted with the test-sample.
Inference criteria: Whether the found model holds up in a test sample, was tested with regard to the four standard measurement invariance assumptions according to Meredith (1993).
3. Hypothesis: We expect the final questionnaire to show good discriminant validity.
Planned Analysis: Pearson's correlation coefficients were calculated with other relevant measures.
Assumptions: For item selection we used random sample split to arrive at a training-sample for the algorithm and a test-sample. The correlations were conducted with the test-sample.
Inference criteria: Correlations were evaluated as follows: correlations >0.1–small, >0.3–moderate, and >0.5–strong.

References

- Algner, M., & Lorenz, T. (2022). You're Prettier When You Smile: Construction and Validation of a Questionnaire to Assess Microaggressions Against Women in the Workplace. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.809862
- Behm, T. (2018). SWE-IV-16. Skala zur Erfassung der Informationsverhaltensbezogenen Selbstwirksamkeitserwartung [Verfahrensdokumentation, Fragebogen deutsche und englische Version (SES-IB-16)]. In Leibniz-Institut für Psychologie (ZPID) (Hrsg.), Open Test Archive. Trier: ZPID. https://doi.org/10.23668/psycharchives.4598
- Beißert, H., Köhler, M., Rempel, M., & Beierlein, C. (2015). Deutschsprachige Kurzskala zur Messung des Konstrukts Need for Cognition NFC-K. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis230
- Bless, H., Wänke, M., Bohner, G., Fellhauer, R. F., et al. (1994). Need for Cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben [Need for cognition: A scale measuring engagement and happiness in cognitive tasks]. *Zeitschrift für Sozialpsychologie, 25*(2), 147–154.
- Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education, 53*(4), 1207–1217. https://doi.org/10.1016/j.compedu.2009.06.004
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116
- Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Re- thinking data literacy in the age of disinformation, misinformation, and mal- information. Internet Policy Review, 9(2). https://doi.org/10.14763/2020.2.1481
- Cui, Y., Chen, F., Lutsyk, A., Leighton, J., & Cutumisu, M. (2023). Data liter- acy assessments: A systematic literature review. Assessment in Education: Principles, Policy & Practice, 30, 1–21. https://doi.org/10.1080/0969594X.2023. 2182737
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. https://doi.org/10.1007/BF02294825

- Moshagen, M., & Bader, M. (2023). Package 'semPower' (Version 1.0.0) [Manual]. CRAN. https://github.com/moshagen/semPower
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2014). Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis76
- Remmert, N., Schmidt, K. M. B., Mussel, P., Hagel, M. L., & Eid, M. (2022). The Berlin Misophonia Questionnaire Revised (BMQ-R): Development and validation of a symptom-oriented diagnostical instrument for the measurement of misophonia. *PLoS ONE, 17*(6), e0269428. https://doi.org/10.1371/journal.pone.0269428
- Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02
- Schauffel, N., Schmidt, I., Peiffer, H., & Ellwart, T. (2021). ICT Self-Concept Scale (ICT-SC25). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis308_exz
- Schmalbach B, Irmer J, Schultze M (2019). _ezCutoffs: Fit Measure Cutoffs in SEM_. R package version 1.0.1, https://CRAN.R-project.org/package=ezCutoffs
- Schneider, J., Striebing, C., Hochfeld, K., & Lorenz, T. (2024). Establishing Circularity: Development and Validation of the Circular Work Value Scale (CWVS). *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1296282
- Schultze, M. (2020). stuart: Subtests Using Algorithmic Rummaging Techniques. *R-Package*. Available online at: https://cran.r-project.org/web/packages/stuart/index.html
- Zhang, C., & Conrad, F. (2014). Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127–135. https://doi.org/10.18148/srm/2014.v8i2.5453