Development of a German Instrument for Self-Rated Data Literacy

An Algorithm-based Approach to Scale Development

Leonie Hagitte

2024-08-25

Contents

Abstract

The increasing relevance of competent and critical handling of data in society not only makes it possible to record this competence, but also makes self-perception with regard to this competence increasingly clear. Previous approaches consider this competence primarily against the specific background of individual target groups, jobs or roles (Cui et al., 2023). In addition, only a few explicitly refer to the general population (Carmi et al., 2020; Cui et al., 2023). In view of the various theoretical approaches, there is a need for a uniform definition of data literacy in order to create comparability. Our aim is therefore to derive a holistic definition based on these approaches and to develop a questionnaire for self-perception of one's own data literacy. To this end, the decisive factors for the construct from previous definitions and operationalizations in various disciplines are brought together. Cognitive interviews are conducted iteratively to create and refine the items. The items are then selected using algorithm-based item selection. The facets of data literacy are comprehensively tested for factorial, discriminant, convergent and congruent incremental validity in order to promote a differentiated understanding of the construct. Construct and criterion validity are tested using correlations and hierarchical regression analyses, while cross-validation checks the robustness of the instrument. Based on a cross-sectional online questionnaire study, we found XXX The self-assessment questionnaire promotes a holistic assessment of competence and its perception for further research, for example by comparing self-assessment and actual performance. *Keywords: Data Literacy, Questionnaire Development, Algorithm-Based Item Selection, Genetic Algorithm*

Acknowledgements

1 Background

In the current digital age characterized by an overwhelming influx especially of online information, it is imperative to accurately distinguish credible, well-substantiated news from various forms of rumors, misinformation, and falsehoods (Koltay, 2017; Leighton et al., 2021; Roetzel, 2019). Previous research has examined our capacity to assess the trustworthiness of news sources and the subsequent effects on online behaviors, such as information-sharing practices (e.g. Pennycook et al., 2021; Pennycook & Rand, 2019). Thus, the relevance of data literacy in today's society becomes evident as it serves as a potent tool in navigating the complex data-driven environment (Carmi et al., 2020; e.g.: Cui et al., 2023; Leighton et al., 2021; Ridsdale et al., 2015). In a world characterized by information overload and rapid technological advancements (Koltay, 2017; Leighton et al., 2021; Roetzel, 2019), individuals equipped with strong data literacy skills can discern patterns, critically evaluate information, and make informed decisions (e.g. Chen et al., 2024; Cui et al., 2023). The exploration of citizens' interaction with media and the cultivation of their agency has traditionally started around concepts such as written literacy, media literacy, information literacy, and digital literacy [QUELLE]. In more recent discussions, data literacy has been approaching relevance among discussed competencies regarding what is necessary for agency in the current society (Carmi et al., 2020; Leighton et al., 2021). Deficiency in data literacy not only exposes individuals to various risks and harms on personal, social, physical, and financial levels but also constrains their capacity to actively engage as informed citizens within an evolving, data-driven society (Carmi et al., 2020; Leighton et al., 2021). Thus, data literacy is a competency that is becoming increasingly important to everyone. Research has acknowledged this in recent years, as more and more research is being done in that direction (Chen et al., 2024; Cui et al., 2023). This study aims to complement the current research, with a self rating questionnaire for assessing data literacy among citizens.

Data literacy involves the ability to effectively collect, manage, evaluate, and apply data in a critical manner (Ridsdale et al., 2015). According to Wolff et al. (2016), it means being able to ask and answer everyday questions using both small and large datasets while considering ethical aspects. This includes skills such as selecting, cleaning, analyzing, visualizing, criticizing, and interpreting data, as well as communicating insights from data and using data for various purposes. Frank (2016) distinguish between cognitive skills, like data collection and analysis, and social skills, which involve trusting data while maintaining skepticism. Calzada Prado & Marzal (2013) outline five dimensions of data literacy: understanding data, acquiring data, interpreting and evaluating data, managing

data, and using data. Understanding data includes knowing their types, roles, and significance, while acquiring data involves evaluating and selecting sources. Interpreting and evaluating data encompass understanding different presentation methods and data interpretation. Managing data includes storage, management, and reuse. Using data involves preparation, analysis, communication, and ethical considerations (Calzada Prado & Marzal, 2013). This small comparison already highlights one prominent feature of data literacy - It is a heterogeneous concept (Chen et al., 2024). Every subject or profession seems to hold their own definition or framework of data literacy (Chen et al., 2024; Cui et al., 2023). While that most certainly is good for assessing specific skills (e.g. in an Recruitment test), it limits the generalisability and comparability of data literacy across individuals with different background. It furthermore limits the accuracy of communication about the topic as two people with different background might hold different definitions on data literacy. In the study from Cui et al. (2023) it also becomes apparent, that one group seems to be underrepresented in the research on data literacy: citizens. Despite being the largest demographic group, citizens are often overlooked in favor of specific professions such as researchers, librarians, students, or educators (Cui et al., 2023). Citizens in this case mean people of the general public, that hold no special role or profession, related to data handling or aspects related to data literacy (Wolff et al., 2016). This trend raises questions about the emphasis on certain aspects of data literacy, many of which tend to align more closely with professional roles than with the needs of laypeople (Schüller, 2020). In their Framework Schüller (2020) highlight the different roles in their data literacy framework: Some of the facets or skills regard "data-consumption", whereas the most are skills "data-producers" would have. This is also reflected when taking a look into related concepts. The definition proposed by Wolff et al. (2016) suggests that data literacy shares some common competencies with statistical and information literacies.

Information literacy, often studied in library sciences, overlaps with data literacy in terms of accessing, critically evaluating, and using data sources (Calzada Prado & Marzal, 2013; Shields, 2005). Wolff et al. (2016) also emphasize the importance of the data inquiry process, starting from identifying problems, designing studies, acquiring data, conducting analysis, to drawing data-based conclusions. In comparison, Gould (2017) argued that data literacy is essentially the same as statistical literacy but with additional competencies needed due to the increasing importance of data. These added competencies include understanding who collects the data, how and why data is collected, and understanding data privacy and ownership (Gould, 2017). However, while statistical literacy focuses on quantitative data and basic statistics, data literacy extends to the ability to understand,

access, evaluate, and use arguments and decision-making based on both quantitative and qualitative data (Cui et al., 2023).

The name data literacy suggests that one is talking of some form of capability, skill or ability. In fact, data literacy can be seen as more on the side of competences or proficiencies. The distinction lies in the fact, that many behaviors, incorporated in data literacy go beyond the mere question of whether a person is "able to do it". It includes the question of one's motivation to do it. This stands in opposition to a former trend, where such literacies were understood cognitively [QUELLEN], thus concerning the mere cognitive predisposition of a person. When talking about data literacy as a competency, the emphasis lies on the possibility to do or not do something, assuming that one has the ability to do it. Thus data literacy incorporates several abilities but also certain personal predispositions or convictions [QUELLE].

When discussing data literacy, it's essential to understand the distinctions between the terms data and information and their relationship to one another. The concepts of data and information are foundational in various fields, yet their precise definitions and relationships are often subject to interpretation (e.g. Koltay, 2017; R. Schneider, 2013). In essence, data can be thought of as raw, unprocessed symbols or observations, lacking inherent meaning (Shannon, 1948). Information on one hand is characterized by the amount of surprise data evokes in another person (Shannon, 1948). Bates (2005) on the other hand emphasizes that data evolves into information through interpretation and organization, becoming comprehensible and relevant within a specific context or purpose. Does that mean that data is entropy while information stands in opposition to it?
Not necessarily. Shannon (1948) defines entropy as a measure of uncertainty or disorder in a system. In information theory, entropy is often associated with the amount of unpredictability or randomness in a set of data. However, entropy can also be viewed as a measure of information content within a system (Brillouin, 1953). This perspective suggests that low entropy corresponds to a high concentration of meaningful information. Similarly, the work of Jaynes (1957) highlights the connection between entropy and information, proposing that information can be quantified in terms of the reduction of uncertainty or entropy in a system. So according to those sources, while data serves as the raw material from which information is derived, it's the reduction of entropy through organization and interpretation that gives rise to meaningful information. Thus, rather than viewing data as synonymous with entropy or information, it is more accurate to consider information as emerging from the structured representation of data.

It is structured representation of data, that also according to the framework of

(Schüller, 2020), citizens are primarily concerned with: Citizens holding roles where they mainly consume data or information. Consequently, they may encounter difficulties with tasks or items related to producing facets such as providing or exploiting data. This imbalance could potentially undermine the fairness of tests and questionnaires designed to assess data literacy, particularly, if these assessments prioritize data and statistical literacy over information literacy. Therefore, our objective was to formulate a definition that adequately addresses both domains, ensuring a balanced representation.

1.1 Conceptual Integration

Thus, via conceptual integration we arrived at the following definition: data literacy is the ability to collect, manage, evaluate, and apply data effectively. It involves asking and answering real-world questions from datasets while considering ethical use. Core skills include selecting, cleaning, analyzing, visualizing, presenting, critiquing, and interpreting data, information and their sources (Cui et al., 2023; Ridsdale et al., 2015; Wolff et al., 2016). The construct can be structured in five facets (Comprehension,Evaluation, Integration, Communication & Statistics)(figure 1) that are also very prominent in most definitions in the literature (Cui et al., 2023). We further divided them into "consumer" facets (Comprehension,Evaluation & Integration), which are relevant for nearly every person in society, from citizens up, as well as "producer" facets (Communication & Statistics), which are mainly relevant for people, actively working with data.

1.1.1 Comprehension

This factor encompasses skills related to understanding and critically evaluating data and information. It involves the ability to comprehend various forms of data presentation, detect inconsistencies, interpret data comprehensively, and identify logical fallacies. Individuals with high scores on this factor demonstrate a strong aptitude for processing and making sense of information across different formats, enabling them to draw accurate conclusions and insights.

1.1.2 Evaluation

This factor involves skills related to critically evaluating information sources and discerning between facts and opinions. It encompasses the ability to assess the credibility and reliability of information, considering factors such as the reputation of the source and the context in which the information was presented.

Individuals scoring high on this factor demonstrate awareness of potential biases or vested interests in information sources.

1.1.3 Integration

This factor relates to the ability to integrate data-driven insights into one's worldview and values. It involves actively seeking comprehensive understanding of various topics, engaging with diverse perspectives, and consciously incorporating data-driven insights. Individuals scoring high in this factor adapt their opinions based on new data, prefer evidence-based information, and ensure their values align with reliable data. They engage with information and perspectives that challenge their existing views, showing a willingness to reassess their opinions and positions based on new data.

1.1.4 Communication

This factor revolves around the skill to effectively communicate and present data through various means, including visual formats, verbal explanations, and written descriptions. It requires translating complex data into clear formats, ensuring comprehension by varied audiences. Individuals scoring high in this factor hold the ability to translate data into simple visualizations, present findings confidently, and articulate complex information effectively in written and visual as well as verbal formats.

1.1.5 Statistics

This factor covers skills related to managing and analyzing data effectively. It involves proficiency in organizing and analyzing data using software tools, conducting statistical analyses, and understanding research methodologies. Additionally, it includes conducting interviews or surveys for data collection, performing basic statistical analysis and recognizing trends in graphical representations. Individuals scoring high on this factor exhibit competence in statistical methods, enabling them to effectively analyze data and interpret findings.

1.2 Delineation from other concepts - Convergent Constructs

This definition of data literacy encompasses certain characteristics and behaviors from similar constructs. Examples for those convergent constructs are *critical*
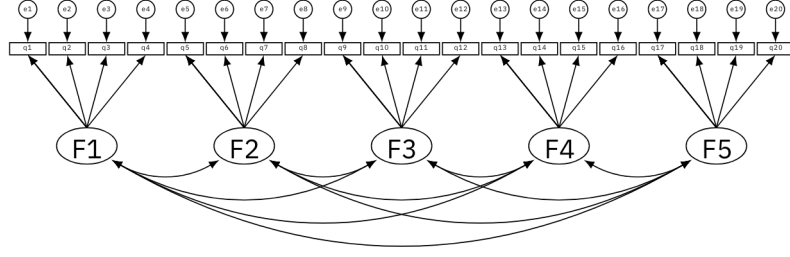
Figure 1: Illustration of the Model data literacy.

thinking, media competency, technology competency, statistical literacy as well as *information literacy* (Chen et al., 2024; Cui et al., 2023; Leighton et al., 2021). As those constructs share substantive parts, differing in size regarding the respective definition of the constructs, it is to be expected that all of them show moderate to strong positive correlations.

Our definition incorporates statistical literacy (Gal, 2002) by emphasizing data interpretation, analysis, and understanding different types of data representations, such as graphs and tables (Statistics). It includes elements of information literacy (Association of College & Research Libraries, 2000) by focusing on evaluating the credibility of data sources, considering factors like reputation and biases, similar to assessing the quality of information sources (Evaluation & Integration)(Webber & Johnston, 2017). Both statistical and information literacy involve using data and information to make informed decisions (Gal, 2002; Webber & Johnston, 2017). Data literacy emphasizes integrating data-driven insights into one's opinions and values, which influence decision-making processes, without focusing on the decision making. Thereby it aligns with the goals of statistical and information literacy (vgl. figure 2). The factor "Comprehension" encompasses behaviors and skills associated with critical thinking (Payan Carreira et al., 2022; Rear, 2019), such as identifying weaknesses in one's reasoning or actively shaping discourse and public dissemination of information. The factors "Evaluation" and "Statistics" encompass behaviors and skills related to technology competency, including navigating and critically evaluating online sources and platforms, using information and communication technology, and utilizing statistical software. The "consumer" factors are closely related to media literacy, focusing on skills related to critically analyzing and interpreting various media formats for understanding and engaging with media content (vgl. figure 2).
In contrast to our definition, statistical literacy often focuses more narrowly on statistical concepts and methods, such as probability, sampling, and hypothesis testing (Gal, 2002). Our definition encompasses a broader range of skills beyond statistical concepts, such as data visualization, software usage, and understand-

ing data collection methods. While information literacy involves assessing the quality of information sources, our definition places a particular emphasis on assessing data quality, considering factors like sample size, biases, and data context. This aspect extends beyond traditional information literacy (Association of College & Research Libraries, 2000) and is more specific to data literacy. Data literacy involves proficiency in using technology, but specifically focuses on understanding and working with data. Technology competency encompasses a broader set of digital skills that extend beyond those relevant for data literacy.
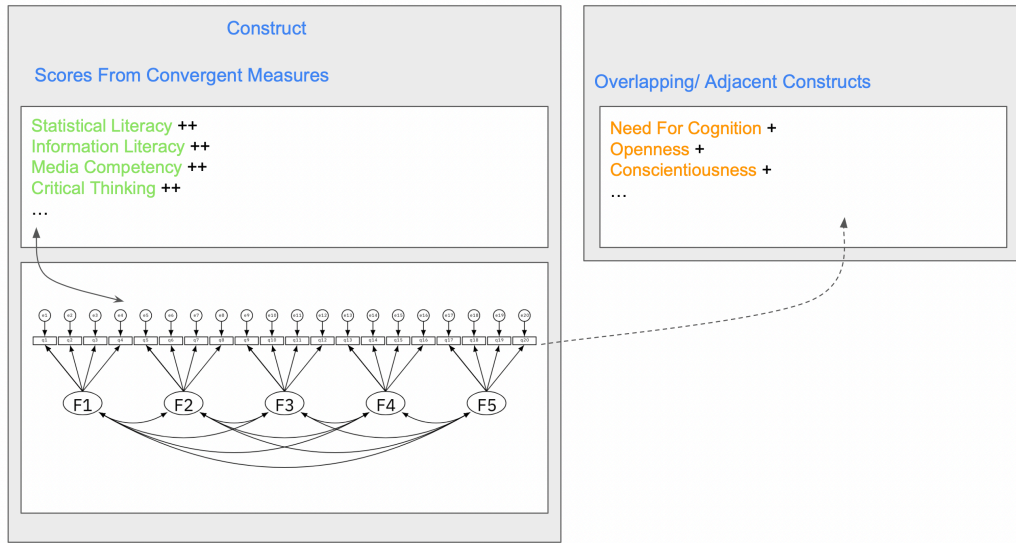

Figure 2: Illustration of the nomological net of data literacy.

1.3 Discrimminant Constructs

1.3.1 Need for Cognition

The personality trait known as *Need for Cognition* (NFC) originated in social psychology during the 1940s and 1950s. The concept of NFC, representing an inclination for joyful thinking, is evident in the works of Maslow (1943), Murphy (1947), Asch (1952), and Sarnoff & Katz (1954). However, the conceptualization of NFC underwent refinement in the mid-1950s through experimental investigations (Cohen et al., 1955). They defined NFC as "a need to structure relevant situations in meaningful, integrated ways. It is a need to understand and make reasonable the experiential world" (Cohen et al., 1955, p. 291). The concept captures individual variations in the engagement and enjoyment of thinking tasks (Bless et al., 1994). As data literacy incorporates several cognitive aspects as well as the motivation to understand data and information one gets presented with, it is expected to correlate positively with one's need for cognition. As need for

cognition is more trait like and data literacy is more a competency and therefore less stable, it should not correlate too highly positive, i.e. correlation should be small (vgl. figure 2).

1.3.2 Opennes to new experiences

Openness to new experiences reflects a broad appreciation for art, emotion, adventure, unconventional ideas, imagination, curiosity, and diverse experiences. Individuals high in openness tend to be intellectually curious, receptive to emotions, appreciative of beauty, and eager to explore new possibilities (John et al., 2008). They are often more creative and emotionally attuned compared to those low in openness. However, they may also be perceived as unpredictable and prone to engaging in risky behaviors, including drug use [@John et al. (2008). High openness is associated with seeking intense and euphoric experiences as a means of self-actualization. In contrast, individuals low in openness tend to seek fulfillment through perseverance and are characterized as pragmatic and sometimes viewed as dogmatic or closed-minded. The interpretation and contextualization of the openness factor remain debated, partly due to a lack of biological evidence supporting this trait. Unlike other personality traits, openness has not shown consistent associations with specific brain regions in neuroimaging studies (DeYoung et al., 2010). As already mentioned, data literacy can be thought of as a competency, thus incorporating the individual motivation, leading to a certain behavior. Therefore, openness to new experiences is expected to correlate positively with data literacy. As openness to new experiences is more trait like and data literacy is not, the correlation is expected to be small (vgl. figure 2).

1.3.3 Conscientiousness

Conscientiousness refers to an individual's propensity for self-discipline, dutifulness, and striving for achievement in alignment with external standards or expectations. It encompasses levels of impulse control, regulation, and goal-directed behavior Toegel & Barsoux (2012). High conscientiousness is characterized by persistence and focus, often perceived as stubbornness, whereas low conscientiousness is linked to flexibility and spontaneity, potentially manifesting as carelessness and unreliability (Toegel & Barsoux, 2012). Individuals with high conscientiousness tend to prefer planned actions over spontaneous ones (Costa & McCrae, 1992; John et al., 2008).
The cognitive aspects of data literacy as well as the motivation to understand data and information speaks to the conscientiousness of people as well. As being

critical and at times detail oriented (e.g. in interpreting results, or spotting incon-
sistencies in presented information or while examining the credibility of sources)
is also integral to data literacy, data literacy and conscientiousness are expected
to correlate positively. As conscientiousness also trait opposing to data literacy,
they should not correlate to highly positive, i.e. correlation should be small (vgl.
figure 2).

1.4 Aim of the Study

The aim was to derive a comprehensive definition of data literacy based on exist-
ing approaches and to develop a questionnaire for self-rated data literacy with cit-
izens being the target population. An emphasis lies on measuring the three con-
sumer factors of the construct (Comprehension,Evaluation & Integration). Thus
the research question of this study is:

"Does the proposed set of items effectively capture the latent factor structure of
self-rated data literacy, and can the created scale be considered a reliable and valid
measure of this construct?"

*H1: The test-data will support the suggested latent factor structure and the proposed
measurement model.*
*H2: The latent factor structure of the initial analysis will be supported by a different
sample.*
*H3: A moderate to high positive correlation with the SWE-IV-16 (Behm, 2018) is ex-
pected.*
*H4: A moderate positive correlation with the the ICT-SC25 (Schauffel, 2021) is ex-
pected.*
H5: A small positive correlation with the NFC-K (Beißert, 2015) is expected.
*H6: A small positive correlation with the the BFI-10 (K. Rammstedt B., 2014) is ex-
pected.*

2 Methods

2.1 Item Creation

A literature review was done to create items and then ten cognitive interviews
were held to refine those potential items. The interviews were administered iter-
atively to refine the items consecutively. The refined items are then asked along-
side other questionnaires in a online survey. We will treat the first 25 participants

like a pilot, to check for potential problems in the survey(like length, spelling mistakes that have been overlooked etc.).

2.2 Sample

The participants are recruited through a combination of personal and professional networks, along with outreach on several online social media platforms (e.g. Instagram, LinkedIn, Whatsapp, Telegram and via e-mail). Conducted in German, the participation in the study was entirely voluntary, with no external incentives provided. 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, 2024) 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; J. Schneider et al., 2024). So the optimal sample size, we are aiming at, lies somewhere between those numbers.

Participants had to be of legal age to be included in the study. Furthermore, 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.

The following characteristics of this study's sample will be made with referral to the respective statistics in the German population of 2022. The sample for this study comprised N = 616 participants. Within the sample, 48,3% identified as female(50,65%), 50,2% as male(49,35%), and 1,5% did not identify with binary gender categories (Statistisches Bundesamt, 2024c). The average age was 40 years (M = 44,6)(Statistisches Bundesamt, 2024b), with an average age of 39 years(M = 38.79; SD = 14.5) amongst women (M = 45,9)(Statistisches Bundesamt, 2023a) and 42 years (M = 41.96; SD = 14.23) amongst men (M = 43,2)(Statistisches Bundesamt, 2023a). The average age of people not identifying with binary gender was 28 years (M = 27.63; SD = 11.0). Regarding education, participants exhibited a [insert educational level- specifying the range or types of educational levels observed in the sample]. Of those participants, who indicated they were in an employment, 64.88% had a full time employment (51,75% of the women indicated full time employment, 77,73% of the men indicated full time employment), at the time (65,15%; Men = 42,40%, Women = 22,75%)(Bundesargentur für Arbeit, 2024; Statistisches Bundesamt, 2023b), 35.12% indicated a part time employment (48,25% of women indicated part time and 22,27% of the men), at the time (28,22%; Men

= 6,16%; Women = 22,06%)(Bundesargentur für Arbeit, 2024; Statistisches Bundesamt, 2023b) and 13.52% had no work at the time of the survey (6,63%)(Bundesargentur für Arbeit, 2024). The study encompassed every sector within the occupational classification at least once (Bundesargentur für Arbeit, 2024). 25.2% of the participants indicated that they were students at the time of the survey (3,39%)(Statistisches Bundesamt, 2024a).

2.3   Open Science Standards

This project uses the reproducibility workflow proposed by Peikert et al. (2021). 'Docker' and 'renv' work together to create a reproducible and portable environment. 'Docker' captures the complete software stack, while 'renv' focuses on managing R package dependencies and providing a clear documentation of the R package environment. This combination ensures that the analysis can be easily reproduced and shared with others in a reliable and transparent manner. It is to be mentioned, that the repro package from Peikert et al. (2021) has slightly changed in its functionality, namely that it does not ensure any longer, that old versions of the used software get reinstalled. Furthermore we used the 'reproducibleRchunks' package from Brandmaier & Peikert (2024). This package enables the verification of computational results in R for reproducibility, ensuring that the same script with the same data produces identical results across different computers or at different times. When knitting the respective document, one can see the results for the respective chunks, as to whether they are reproducible or not. The study was preregistered at Zenodo (DOI:10.5281/zenodo.11196495).

2.4   Procedure

A cross-sectional online survey is used to examine a sample from the general population. Participants complete the Self-perceived Data Literacy Scale alongside demographic questions and additional validation measures. Survey questions of each measurement are randomized for each participant to minimize order effects and response biases. To shorten the overall length of the assessment 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.

2.5 Instruments

2.5.1 Measuring Data Literacy

On Data Literacy the participants will be 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.

2.5.2 Measuring Information Literacy

The SWE-IV-16 (Behm, 2018) (McDonalds ω = .91; Cronbachs α =.91) 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). In our study this construct is used as a proxy for information literacy. 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"). 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. The final questionnaire is expected to correlate moderately up to highly positive with the SWE-IV-16 (Behm, 2018), measuring peoples ability to engage in information behaviour.

2.5.3 Measuring Need for Cognition

The NFC-K (Beißert, 2015) (McDonalds ω = .62; Cronbachs α =.60) 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 & Petty (1982) and translated by Bless et al. (1994). 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. A small to moderate positive correlation with the NFC-K (Beißert, 2015) is expected, assessing the Need for Cognition (NFC).

2.5.4 Measuring Technology Competency

To assess self-perceived competence in using information and communication technology (ICT), the five general items of the ICT-SC25 (Schauffel, 2021) will be used (McDonalds ω = .93; Cronbachs α =.93). 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). 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, 2021). A moderate, positive correlation of the final scale with the five general items of the ICT-SC25 (Schauffel, 2021) is expected.

2.5.5 Measuring Openness and Conscientiousness

The BFI-10 (K. Rammstedt B., 2014) will be used to assess personality based on the five-factor model. Only the items on openness (McDonalds ω = .63; Cronbachs α =.63) and conscientiousness (McDonalds ω = .56; Cronbachs α =.56) were assessed. The items are answered on a five-point rating scale from "strongly disagree" (1) to "strongly agree" (5). To measure the respondent's individual traits on the two 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 B. Rammstedt & John (2007) for reference values). Small to moderate positive correlations with openness and conscientiousness of the BFI-10 (K. Rammstedt B., 2014) are expected.

3 Analysis

3.1 Data Quality

Careless or inattentive response patterns were analyzed by the ATC items. Furthermore, the data was checked for outliers. Because of the planned random missings in the data, those missings were imputed via the 'mice' package, with predictive mean matching. Also for model estimation, full information maximum likelihood was used to handle the missing data.

3.2 Main Analyses

Algorithm based item selection via the R package 'stuart'(Schultze, 2022) and CFA via the R package 'lavaan'(Rosseel, 2012). Inference criteria: Model fit will be assessed using established criteria (e.g.: Hu & Bentler (1999)). Comprising of Chi square 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)

3.2.1 Rationale for Measurement Model

One decision that needs to be done by the researcher before the item selection with 'stuart', is the design of the measurement model, or how many items per factor the final scale should have. This scale will be created as parsimonious as possible, while ensuring that there is a real model fit to be estimated. So the decision in this case was of statistical nature, while with three items per factor, the model would have been just identified, four items per factor is the most parsimonious choice, where there is already a real model fit, that can be assessed in the end.

3.2.2 Meta-Heuristics

Algorithm-based item selection is used to choose the most relevant items, reducing the initial item pool. In classical approaches, items are evaluated within the overall item pool and are then often selected based on their individual properties (e.g. difficulty, discrimination, item-scale-correlations). Compared to classical approaches, algorithms are more objective and efficient in finding a good or nearly perfect solution with regard to certain criteria (Leite et al., 2008; Olaru et al., 2015). Furthermore, some empirical studies suggest that the use of algorithms leads to

similar or better results in scale construction than traditional approaches (Olaru & Danner, 2021; Sandy et al., 2014; Schroeders, 2016). The automated approach takes the opposite perspective to the classical approach, the one of meta heuristics, by repeatedly estimating CFAs for a multitude of possible item-combinations (Schultze, 2017). Thus, a pool of items with some constraints and the goal to find the one combination that best fits the suggested purpose (e.g. equation 1) of the final scale is estimated (Schultze, 2017).Thus, the selection of items and construction of a questionnaire can be viewed as a combinatorial problem, like the knapsack problem ("Choose a set of objects, each having a specific weight and monetary value, so that the value is maximized and the total weight does not exceed a predetermined limit")(Kerber et al., 2022; Schroeders, 2016, p. 4; Schultze, 2017).

$$\Phi = F(\text{RMSEA}) + F(CFI) + F(\text{SRMR}) \ (1)$$

'Stuart'(Schultze, 2022) can construct subsets from a pool of items by using ant-colony-optimization, genetic algorithms, brute force, or random sampling (Schultze, 2017). Those meta-heuristics like are utilized to handle these combinatorial optimization problems. For this study, a set of 20 items from a set of 71 items is selected, to form a questionnaire. The data literacy self rating scale is optimized for model fit criteria (RMSEA, SRMR, CFI). Those criteria will be defined in the objective function (equations 1 & 2) in 'stuart'Schultze (2017).

We use the genetic algorithm of 'stuart' for the item selection. Genetic algorithms are based on Darwinian evolution principles – selection, crossover, mutation and survival of the fittest [Holland (1992); Schroeders (2016)). With the genetic algorithm, the initial set of 71 items is to be reduced based on the evolutionary process of selection, but opposing to evolution with a goal: A near-optimal "solution". The survival of an item is determined by its quality (called "fitness")(Galán et al., 2013). The algorithm is build on two processes: Variation (i.e. recombination and mutation) and selection. Variation rewards diversity and innovation of items, whereas selection rewards quality or fitness. The algorithm links "genes" (i.e. items), that represent a certain variable, to a "chromosome" (i.e. a set of items). A predefined number of chromosomes are randomly generated from the 1st generation (i.e. the original item pool). The algorithm tries to maximize the psychometric quality of the "chromosomes"(i.e. item sets) by evaluating the "chromosomes"(i.e. item sets) against a "fitness"function. Based on this fitness function, the fittest "chromosomes"(i.e. item sets) of each generation are determined, which then form the basis for the next generation (forming the selection process). The process of variation establishes genetic diversity and mutation within the generations by spontaneously exchanging items within a scale or between two scales, which adds a

degree of randomness to the selection process. This is done a predefined number of iterations. Thereby, the fittest "chromosome"(i.e. set of items), with the highest quality, is to be identified (Schroeders, 2016).

3.3 Validation

The provided sample is divided into two subsets (i.e. training data and test data) using the 'holdout' function in 'stuart'. The specified item-selection procedure is applied to the training data. The training data is undergoing k-fold cross validation (k=3), using the 'kfold' function in 'stuart'. Those kfolded selections are then again iterated three times, to enhance the stability of the solution even more. The other sample, the test data, is used for evaluation of the final models performance, as well as the latent correlations with the convergent measures. Validation with the test data is conducted (using an MG-CFA in 'lavaan') to assess the invariance of the measurement models between the training and testing datasets. Invariance levels are assessed using the criteria of F. F. Chen (2007). Invariance is necessary to claim that the scale validation has worked.

3.4 Exploratory Analyses

Because the complexity of the model, as well as the several objectives of the initially planned objective function, the genetic algorithm had problems converging into a solution, let alone a solution with sufficient model fit. Therefore some objectives were dropped (the composite reliability and the varying item intercepts per factor), in favor of model fit. Different solutions were systematically explored, dependent on the estimator (MLR or WLSMV) and the respective data structure (data treated as metric or ordinal data), as well as on the objective function, to lead to the best solution. That way four models got selected by the algorithm. And a fifth model was created based on the best model of those four models, but with slight "manual" changes; exchanging some items, that showed very low loadings, for others with higher loadings of the respective latent factor.

4 Results

4.1 Model Fit and Measurement Invariance

Of the four explored conditions and respective models, the following model is to be highlighted. An overview over all models and their model fit can be found

in the supplementary material. With an objective function, only optimising for model fit criteria (RMSEA, SRMR, CFI), the algorithm selected 20 of the 71 original items representing the five factors Comprehension, Evaluation, Integration, Communication and Statistics with four items each (Figure X). The solution exhibits good model fit (according to Hu & Bentler, 1999) with ordered data and WLSMV as estimator: Satorra-Bentler-X^2 (df = 414, N = 373) = 582.582, p < 0.001, CFI = .96, TLI = .96, SRMR = .08, RMSEA = .05, 90%-CIRMSEA [.036; .054]. Standardized loadings of the factor Comprehension ranged from .09 to .65. For the factor Evaluation loadings ranged from .44 to .64. For the factor Integration loadings ranged from .35 to .72. For the factor Communication loadings ranged from .75 to .82. For the factor Statistics loadings ranged from .41 to .68. The complete model is also displayed in figure X. All factor loadings including standard errors can be found in the Supplementary Material. Cross-validation of the MG-CFA with the test-data indicated that the assumption of scalar invariance holds across the two subsamples: X^2(df = 389, N = 373) = 659.746, p < 0.001, CFI = .94, SRMR = .069, RMSEA = .053; ΔCFI = 0.003 , ΔSRMR = .003, ΔRMSEA = .003.

Due to the really low loadings of some of the factors onto some respective items. Those items were exchanged manually to explore the resulting impact on the overall model fit, as well as on the MI testing. The manually adapted solution exhibits good model fit (according to Hu & Bentler, 1999) with ordered data and WLSMV as estimator: Satorra-Bentler-X^2 (df = XXX, N = 373) = XXX, p < XXX, CFI = XXX, TLI = .XXX, SRMR = XXX, RMSEA = .XXX, 90%-CIRMSEA [.XXX; .XXX]. Standardized loadings of the factor Comprehension ranged from .XXX to XXX. For the factor Evaluation loadings ranged from .XXX to XXX. For the factor Integration loadings ranged from XXX to .XXX. For the factor Communication loadings ranged from .XXX to .XXX. For the factor Statistics loadings ranged from XXX to XXX. The complete model is also displayed in figure X. All factor loadings including standard errors can be found in the Supplementary Material. Cross-validation of the MG-CFA with the test-data indicated that the assumption of scalar invariance holds across the two subsamples: X^2(df = XXX, N = XXX) = XXX, p < XXX, CFI = XXX, SRMR = XXX, RMSEA = XXX; ΔCFI = .XXX, ΔSRMR = .XXX, ΔRMSEA = XXX.

4.2   Latent Correlations

For the latent correlations, Kendall's rank correlation coefficient - Kendalls tau was estimated, because the data was not normally distributed. The factors (Comprehension, Evaluation, Integration, Communication & Statistics) of the data literacy scale correlated moderately to highly with the SWE-IV-16(τ =.36, τ =.44, τ
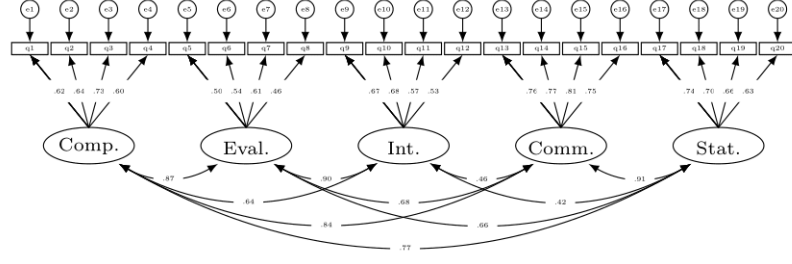
Figure 3: Measurement Model of Model 5.

=.36, τ =.50, τ =.43 ; p < .01). The factors of data literacy showed small to moderate correlations with the NFC-K(τ =.17, p < .01; τ =.24, p < .01; τ =.27, p < .01; τ =.32, p < .01; τ =.36, p < .01). The factors of the data literacy scale showed moderate correlations with the general items of the ICT-SC25(τ =.26, τ =.20, τ =.19, τ =.41, τ =.28 , p < .01). The factors Evaluation, Integration and Communication correlated slightly negative with the openness of the BFI-10(τ = -.14, p < .01; τ = -.10, p < .05; τ = -.11, p < .05). Openness did not correlate statistically significant with the other factors. The factors of the data literacy scale correlated slightly positive with conscientiousness of the BFI-10(τ = .14, p < .01; τ = .13, p < .01; τ = .17, p < .01; τ = .16, p < .01; τ = .18, p < .01). The latent correlations with respective confidence intervalls are also displayed in table 3.

4.3   Reliability and Sensitivity Analysis

The model of the final set of items shows McDonald's ω total = .92. The composite values McDonald's ω total of the factors are Comprehension = .85, Evaluation = .70, Integration = .65, Communication = .68 and Statistics = .56.

residuals correlates or the residuals for the adjacent constructs

5   Discussion

This study meant to examine whether a pool 20 items out of the initial item pool of 71 items, would reflect the suggested measurement model within the current sample. Furthermore, the aim was to find a solution via 'stuart' that yields good model fit as well as reliability and that shows measurement invariance with a random split-sample. Additionally, latent correlations were to be examined to locate the construct in the nomological net.

The final model showed acceptable fit (XXX), according to the combinational rule for RMSEA and SRMR of Hu & Bentler (1999). But not according the rules

that are most commonly used not to the combinational rule, that resulted in the least sum of type 1 and type 2 error rates (Hu & Bentler, 1999). Comparison of the training sample and the test sample on the chosen model, measurement invariance analyses indicated

5.1 Summarising the results

Everything worked - what do I need to report? First, report on the three core decisions and how the solution was created, whether it is stable, and how it was validated. The final variant is then reported exactly like a CFA (e.g. Jackson et al., 2009). To have a clear structure, answer the following questions: How large was the original item pool and how was it created? What is the structure of the scale? Which algorithm was used ? Which objective function was used? How stable is the final variant? How was the final variant further validated (e.g. crossvalidate, k-fold)?

The xxx correlations (r = , p) between the newly created scale and the SWE-IV-16, NFC-K, the general items of the ICT-SC25 and openness and conscientiousness of the BFI-10,respectively, indicate the data literacy scale measures #a similar, yet distinct concept#.

5.2 What does it all mean / "Why?"

- Start with the research question

  - Maybe then towards the hypotheses

- connecting findings to the related theories
- very related/ current literature first, than broader is possible
- When discussing the why - be careful, because you didnt test that

5.3 Limitations

- model fit
- ceiling effects
- melted answer categories
- content validity
- reliability!
- DIF?
- psych science - authors guide to generelizability

- attempts to control for limitating factors
- dont include to general/ broad critiques, but special one for my own study

The results of this study should be interpreted with several limitations in mind. The sample deviates from the general population in multiple demographic variables, potentially compromising its representativeness and generalizability. Occupational distribution among participants shows clustering in fields such as *"Gesundheit, Soziales, Lehre und Erziehung"*, *"Buchhaltung, Recht und Verwaltung"*, *"Kaufmännische Dienstleistungen, Vertrieb, Tourismus"* and especially *"Naturwissenschaft, Geografie und Informatik"*. This indicates a selection bias, likely due to recruitment methods (who is reached) and implicitly favoring individuals more interested in data literacy.

The item pool for the questionnaire was specifically trained on this non-representative sample, which may affect its validity. The heterogeneous nature of the construct complicates global instrument development and understanding across all participants. The measure was designed for citizens, potentially limiting discrimination at higher item difficulties or among more literate participants, a direction to be improved in future studies. Also, as data literacy is a heterogeneous construct, the questionnaire could incorporate more aspects to better reflect its full scope, thereby increasing content validity. Expanding the questionnaire with additional items could address this need, although it would deviate from the principle of parsimony.

Additionally, the training and testing data sets differed in size, which could influence measurement invariance testing (F. F. Chen, 2007). While the sample sizes were appropriate, they were at the lower threshold of the prior power analysis (Hu & Bentler, 1999; e.g., Kass & Tinsley, 1979), suggesting that larger samples might have been better. Lastly, although the questionnaire showed good model fit, it should be noted that algorithm-based item selection is a heuristic approach, rather than deterministic, and may not always yield the optimal solution(Blum & Roli, 2003; Schultze, 2017).

5.4   Future directions

Future research should explore adaptive testing using Item Response Theory (IRT). IRT provides a method to tailor item difficulty to respondents' ability levels in real-time, enhancing assessment efficiency and precision. This reduces the number of items required while maintaining high measurement accuracy. Implementing IRT is particularly advantageous for heterogeneous constructs like data literacy, as it ensures each participant is evaluated with items suited to their

skill level. One of the significant challenges in applying IRT is the assumption of unidimensionality, where items are presumed to measure a single underlying trait. Data literacy, however, is a multi-faceted construct, and future studies should investigate the dimensionality of the scale rigorously. An alternative to IRT-based adaptive testing is the use of Classification and Regression Trees (CART). CART is a tree-based method that splits data into subsets based on binary decisions, optimizing for predictive accuracy. This approach could simplify adaptive testing by using binary splits to classify respondents into different levels of data literacy. The Gini index can be employed within CART to identify the optimal cutoff points for these splits, ensuring that each branch of the tree maximally distinguishes between different levels of data literacy competence.

Construct validity is evaluated through confirmatory factor analysis (CFA), using MLR as estimator, as well as correlation analyses with related constructs

- what to optimize the scale for?
- dynamic fit indices
- residuals correlates
- is NOT optimized for model fit criteria and composite reliability (RMSEA, SRMR, CFI & McDonalds ω) as well as variability in the difficulty of items.

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

Asch, S. (1952). *Social psychology.* Prentice Hall.

Association of College & Research Libraries. (2000). *Information literacy competency standards for higher education.* Brochure; American Library Association.

Bates, M. J. (2005). An introduction to metatheories, theories, and models. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences* (2nd ed., pp. 109–121). Taylor & Francis.

Behm, T. (2018). *SWE-IV-16: Skala zur erfassung der informationsverhaltensbezogenen selbstwirksamkeitserwartung [verfahrensdokumentation, fragebogen deutsche und englische version (SES-IB-16)]* [Open Test Archive]. Leibniz-Institut für Psychologie (ZPID). https://doi.org/10.23668/psycharchives.4598

Beißert, K., H. (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., & Schwarz, N. (1994). Need for cognition: Eine skala zur erfassung von engagement und freude bei denkaufgaben [presentation and validation of a german version of the need for cognition scale]. *Zeitschrift Für Sozialpsychologie, 25*, 147–154.

Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Comput. Surv., 35*(3), 268–308. https://doi.org/10.1145/937503.937505

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/https://doi.org/10.1016/j.compedu.2009.06.004

Brandmaier, A. M., & Peikert, A. (2024). Automated reproducibility testing in r markdown. *Preprint*.

Brillouin, L. (1953). Negentropy principle of information. *Journal of Applied Physics, 24*(9), 1152–1163.

Bundesargentur für Arbeit. (2024). *Arbeitslosenzahl in deutschland im jahresdurchschnitt von 2005 bis 2024*. https://de.statista.com/statistik/daten/studie/1223/umfrage/arbeitslosenzahl-in-deutschland-jahresdurchschnittswerte/#:~:text=Im%20Monat%20Juni%202024%20waren,um%20rund%20178.800%20Personen%20höher; Statista.

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

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri, 63*(2), 123–134. https://doi.org/10.1515/libri-2013-0010

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review, 9*(2). https://doi.org/10.14763/2020.2.1481

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, Cui, Y., Lutsyk-King, A., Gao, Y., Liu, X., Cutumisu, M., & Leighton, J. P. (2024). Validating a novel digital performance-based assessment of data literacy: Psychometric and eye-tracking analyses. *Education and Information Technologies, 29*(8), 9417–9444. https://doi.org/10.1007/s10639-023-12177-7

Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation

of need for cognition. *The Journal of Abnormal and Social Psychology*, *51*(2), 291–294. https://doi.org/10.1037/h0042761

Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, *6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

Cui, Y., Chen, F., Lutsyk, A., Leighton, J., & Cutumisu, M. (2023). Data literacy assessments: A systematic literature review. *Assessment in Education: Principles, Policy & Practice*, *30*, 1–21. https://doi.org/10.1080/0969594X.2023.2182737

DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., & Gray, J. R. (2010). Testing predictions from personality neuroscience: Brain structure and the big five. *Psychological Science*, *21*(6), 820–828. https://doi.org/10.1177/0956797610370159

Frank, M. (2016). Data literacy - what is it and how can we make it happen? *Journal of Community Informatics*, *12*, 4–8.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review / Revue Internationale de Statistique*, *70*(1), 1–25. http://www.jstor.org/stable/1403713

Galán, S. F., Mengshoel, O. J., & Pinter, R. (2013). A novel mating approach for genetic algorithms. *Evolutionary Computation*, *21*(2), 197–229. https://doi.org/10.1162/EVCO_a_00067

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22–25. https://doi.org/10.52041/serj.v16i1.209

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence.* The MIT Press. https://doi.org/10.7551/mitpress/1090.001.0001

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620–630.

John, O., Naumann, L., & Soto, C. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In *Handbook of Personality: Theory and Research, 3 Edn.* (pp. 114–158).

Kass, R. A., & Tinsley, H. E. A. (1979). Factor analysis. *Journal of Leisure Research*, *11*(2), 120–138. https://doi.org/10.1080/00222216.1979.11969385

Kerber, A., Schultze, M., Müller, S., Rühling, R. M., Wright, A. G. C., Spitzer, C., Krueger, R. F., Knaevelsrud, C., & Zimmermann, J. (2022). Development of

a short and ICD-11 compatible measure for DSM-5 maladaptive personality traits using ant colony optimization algorithms. *Assessment*, *29*(3), 467–487. https://doi.org/10.1177/1073191120971848

Koltay, T. (2017). Information overload in a data-intensive world. In A. J. Schuster (Ed.), *Understanding information: From the big bang to big data* (pp. 197–217). Springer International Publishing. https://doi.org/10.1007/978-3-319-59090-5_10

Leighton, J. P., Cui, Y., & Cutumisu, M. (2021). Key information processes for thinking critically in data-rich environments. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.561847

Leite, W. L., Huang, I.-C., & Marcoulides, G. A. (2008). Item selection for the development of short forms of scales using an ant colony optimization algorithm. *Multivariate Behavioral Research*, *43*, 411–431. https://doi.org/10.1080/00273170802285743

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370–396. https://doi.org/10.1037/h0054346

Moshagen, M., & Bader, M. (2024). semPower: General power analysis for structural equation models. *Behavior Research Methods*, *56*, 2901–2922. https://doi.org/10.3758/s13428-023-02254-7

Murphy, G. (1947). *Personality: A biosocial approach to origins and structure.* Harper.

Olaru, G., & Danner, D. (2021). Developing cross-cultural short scales using ant colony optimization. *Assessment*, *28*(1), 199–210. https://doi.org/10.1177/1073191120918026

Olaru, G., Witthöft, M., & Wilhelm, O. (2015). Methods matter: Testing competing models for designing short-scale big five assessments. *Journal of Research in Personality*, *59*, 56–68. https://doi.org/10.1016/j.jrp.2015.09.001

Payan Carreira, R., Sacau-Fontenla, A., Rebelo, H., Sebastião, L., & Pnevmatikos, D. (2022). Development and validation of a critical thinking assessment-scale short form. *Education Sciences*, *12*, 938. https://doi.org/10.3390/educsci12120938

Peikert, A., Van Lissa, C. J., & Brandmaier, A. M. (2021). *Reproducible research in r: A tutorial on how to do the same thing more than once.* https://doi.org/10.31234/osf.io/fwxs4

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reason-

ing. *Cognition, 188,* 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

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, K., B. (2014). Big five inventory (BFI-10). *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen (ZIS).* https://doi.org/10.6102/zis76

Rear, D. (2019). One size fits all? The limitations of standardised assessment in critical thinking. *Assessment & Evaluation in Higher Education, 44,* 664–675.

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,* 1–27. https://doi.org/10.1371/journal.pone.0269428

Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B., & Ali-Hassan, H. (2015). *Strategies and best practices for data literacy education knowledge synthesis report.* https://doi.org/10.13140/RG.2.1.1922.5044

Roetzel, P. G. (2019). Information overload in the information age: A review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Business Research, 12*(2), 479–522. https://doi.org/10.1007/s40685-018-0069-z

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

Sandy, C. J., Gosling, S. D., & Koelkebeck, T. (2014). Psychometric comparison of automated versus rational methods of scale abbreviation: An illustration using a brief measure of values. *Journal of Individual Differences, 35,* 221–235. https://doi.org/10.1027/1614-0001/a000144

Sarnoff, I., & Katz, D. (1954). The motivational bases of attitude change. *The Journal of Abnormal and Social Psychology, 49*(1), 115–124. https://doi.org/10.1037/h0057453

Schauffel, S., N. (2021). ICT self-concept scale (ICT-SC25). *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen (ZIS).* https://doi.org/10.6102/zis308_exz

Schmalbach, B., Irmer, J. P., & Schultze, M. (2019). *ezCutoffs: Fit measure cutoffs in SEM.*

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

Schneider, R. (2013). Research data literacy. In S. Kurbanoğlu, E. Grassian, D. Mizrachi, R. Catts, & S. Špiranec (Eds.), *Worldwide commonalities and challenges in information literacy research and practice* (pp. 134–140). Springer International Publishing.

Schroeders, O. A. O., Ulrich AND Wilhelm. (2016). Meta-heuristics in short scale construction: Ant colony optimization and genetic algorithm. *PLOS ONE*, *11*(11), 1–19. https://doi.org/10.1371/journal.pone.0167110

Schüller, K. (2020). *Future skills: A framework for data literacy* (Working Paper No. 53). Hochschulforum Digitalisierung. https://doi.org/10.5281/zenodo.3946067

Schultze, M. (2017). *Constructing subtests using ant colony optimization* [Doctoral dissertation]. Freie Universität Berlin.

Schultze, M. (2022). *Stuart: Subtests using algorithmic rummaging techniques.*

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Shields, M. (2005). Information literacy, statistical literacy, data literacy. *IASSIST Quarterly*, *28*(2–3), 6. https://doi.org/10.29173/iq790

Statistisches Bundesamt. (2023a). *Durchschnittsalter der bevölkerung in deutschland nach geschlecht von 2011 bis 2022.* https://de.statista.com/statistik/daten/studie/1084446/umfrage/durchschnittsalter-der-bevoelkerung-in-deutschland-nach-geschlecht/; Statista.

Statistisches Bundesamt. (2023b). *Leichter rückgang: Vollzeitbeschäftigte arbeiteten 2022 durchschnittlich 40,0 wochenstunden.* [Pressemitteilung Nr. N047 vom 28. August 2023]. https://www.destatis.de/DE/Presse/Pressemitteilungen/2023/08/PD23_N047_13.html

Statistisches Bundesamt. (2024a). *Anzahl der studierenden an hochschulen in deutschland in den wintersemestern von 2002/2003 bis 2023/2024.* https://de.statista.com/statistik/daten/studie/221/umfrage/anzahl-der-studenten-an-deutschen-hochschulen/; Statista.

Statistisches Bundesamt. (2024b). *Bevölkerung nach dem gebietsstand und durchschnittsalter 1990 bis 2023.* https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/bevoelkerungsstand-gebietsstand-werte.html.

Statistisches Bundesamt. (2024c). *Bevölkerung nach nationalität und geschlecht 1970 bis 2023 in deutschland.* https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/deutsche-nichtdeutsche-bevoelkerung-nach-geschlecht-deutschland.

Toegel, G., & Barsoux, J.-L. (2012). How to become a better leader. *MIT Sloan Management Review*, *53*, 51–60.

Webber, S. A., & Johnston, B. (2017). Information literacy: Conceptions, context and the formation of a discipline. *Journal of Information Literacy, 11.* https://doi.org/10.11645/11.1.2205

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics, 12*(3), 9–26. https://doi.org/10.15353/joci.v12i3.3275