Digital Trace Data as Measurement Instruments for Variance-Theoretic Research in Information Systems

Joschka Andreas Hüllmann[1]

[1] University of Twente, Enschede, The Netherlands
j.huellmann@utwente.nl

AUTHOR ACCEPTED MANUSCRIPT. Cite as:

Hüllmann, J.A. (2026). Digital Trace Data as Measurement Instruments for Variance-Theoretic Research in Information Systems. In: Wurm, B., Mendling, J. (eds) *Digital Trace Data Research in Information Systems*. Progress in IS. Springer, Cham. https://doi.org/10.1007/978-3-032-05497-5_10

Abstract. Driven by the digitization of organizations, digital trace data offer novel insights into human behaviors with technology. Digital trace data are longitudinal records of technology use. Over the last years, we have seen a surge in interest with growing empirical applications and research into the conceptual and methodological foundations of digital trace data research. So far, however, using digital trace data as a basis for measurement instruments in traditional variance-theoretical applications has received little attention, although they may enable novel analyses for theorizing from digitized contexts. The nascent research using digital trace data as measurement instruments has received critiques about validity problems, suggesting that guidelines for robust construct operationalizations are needed. Based on a literature review, this chapter identifies sources for validity problems with digital trace data. I further derive recommendations for assessing and reporting instrument validity with digital trace data. Thereby, this chapter contributes to improving the robustness of quantitative research using digital trace data.

Keywords: Digital trace data, instrument development, construct validity, variance theory.

1 Introduction

Digital trace data research occurred in 16% of the papers in the basket of eight journals in 2018 (Grover et al., 2020). Digital trace data are historical, longitudinal logs of human behaviors and actions that are generated through technology use and promise new research opportunities (Hüllmann, 2019). Empirical applications range from analyzing social media traces to app usage or sensor data that capture heart rate variability. Despite the widespread availability of digital trace data, the predictability of human behavior with digital trace data is a highly contested question. While Narayanan puts behavioral research using digital trace data down as "[…] essentially an elaborate random number generator" (Narayanan, 2019), others suggest that the social behavior of humans is predictable to a high degree given the necessary data (Pentland, 2015; Song et al., 2010; Stewart, 2019).

Research into the conceptual and methodological foundations of digital trace data and computational methods is surging in the times of big data, algorithms, and machine learning (e.g., Braun & Kuljanin, 2015; Rothmeier et al., 2021; Xu et al., 2020). Less, but to an increasing extent, people examine how to theorize from digital trace data (e.g., Berente et al., 2018; Miranda et al., 2022; Pentland et al., 2020, 2021). Until now, the information systems (IS) field has not much engaged in a deeper reflection of digital trace data as a basis for measurement instruments. The nascent conversation hardly addresses constructs and measurements for variance-theoretic models, although different types of digital trace data exist that can operationalize theoretical constructs. To address this gap, this chapter discusses digital trace data-based instruments for construct measurement, which enable novel analyses for theorizing from digitized contexts (Pentland et al., 2021).

The focus on digital trace data for building measurement instruments is important because they promise more objectivity, larger scale data sets, and novel inferences over traditional survey-based methods and established scales. However, existing research using digital trace data for operationalizing theoretical constructs has been criticized. The critique deplores, inter alia, opaque assumptions about how theory is operationalized, resulting in questionable instrument validity (Howison et al., 2011; Johnson et al., 2019). Previous research has neglected assessing and reporting instrument validity and making assumptions about measurements explicit. This chapter summarizes this critique and proposes guidelines for overcoming these pitfalls in upcoming studies. Thus, I pose the research question: :

How can digital trace data be used as robust instruments for construct measurement?

This chapter addresses the research question through a literature review, following Schryen (2015). First, I identify the problems that occur when using digital trace data as measurement instruments. Then, I relate these problems to the best practices from measurement theory that tackle comparable problems in traditional instrument development. Applying these practices to building digital trace data-based instruments, I propose a starting point for conducting variance-theoretic research using digital trace data. Thereby, this chapter contributes methodological insights into how digital trace data can be used as construct measurements. It encourages digital trace data for variance-theoretic applications and provides recommendations for improving instrument validity (cf. MacKenzie et al., 2011). These recommendations help researchers to ensure robust operationalizations using digital trace data.

2 Background

This section introduces the concept of variance theory and shows how digital trace data relate to it. The subsequent need for instrument validity and rigorous instrument development when using digital trace data for variance-theoretic research are motivated.

2.1 Variance and process theory

Variance theories explain relationships between theoretical constructs by understanding the variation among measurements for the related constructs within a nomological network (Burton-Jones et al., 2015). A nomological network is a network representation of the interrelated theoretical constructs from a narrow part of a theoretical domain (e.g., antecedents, focal construct, mediators, outcomes). In variance-theoretic studies, researchers seek to understand how the related constructs contribute to the observed variation in the focal construct of interest, i.e., how an independent variable is related to change in a dependent variable. For example, an increase in *perceived usefulness* (independent variable) is related to an increase in *technology use* (dependent variable). Variance-theoretic approaches are well-suited for theory testing in hypothetico-deductive settings (i.e., hypotheses are derived from the literature and empirically tested). By understanding the sources of variance and their underlying mechanisms, researchers develop more nuanced theories and interventions. Typical data sources include archival data or surveys, and common methods to understand the sources of variance include the analysis of variance (ANOVA), regression analysis, or structural equation modeling (SEM) (Recker, 2021, chapter 5).

An alternative to variance theories are process theories. Process theories seek to "provide explanations in terms of the sequence of events leading to an outcome […] by understanding patterns in events" (Langley, 1999, p. 692). Temporality plays a crucial role in process theorizing, and patterns can comprise events, activities, and choices along the temporal dimension (Figure 1). Naturally, digital trace data facilitate the application of computational methods and following process-theoretic approaches, such as process mining, network analyses, or qualitative, interpretive analyses, because digital trace data are longitudinal logs of human behaviors and actions (Miranda et al., 2022; Hüllmann & Krebber, 2020).

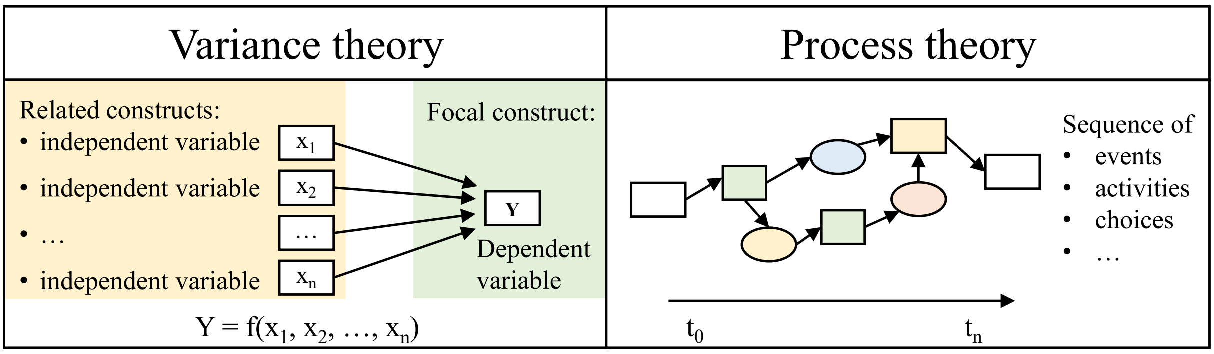

Figure 1. Comparison of variance and process theory (adapted from Langley, 1999)

Digital trace data also promise novel insights for variance-theoretic approaches because the richness of the data goes beyond traditional survey-based methods and archival data. Due to the characteristics of digital trace data, however, special caution must be taken when implementing variance-theoretic approaches such as regression analyses to ensure instrument validity (Olteanu et al., 2019).

2.2 Instrument validity

Linking theoretical constructs to measurement instruments is called *operationalization*, and instrument validity describes the *quality* of an operationalization (Braun & Kuljanin, 2015). Instrument validity severely impacts making robust inferences and has been the subject of extended discussions in information systems and adjacent disciplines such as psychology (Messick, 1995). Cortina et al. (2020, p. 1351) report that "many of the most influential articles of the first decade of the Journal of Applied Psychology (JAP) related to measurement in some way". Since then, it has been dubbed one of the most important concerns in measurement theory (Westen & Rosenthal, 2003), and it is experiencing a rejuvenescence due to the open science reform. Researchers are encouraged to assess and report the validity of their measurements meticulously to ensure replicability (Flake, 2021; Flake & Fried, 2020).

Instrument validity is a higher-level concern that comprises construct, content, and criterion validity. Construct validity describes how well a measurement instrument substantively measures a latent theoretical construct of interest (Cronbach & Meehl, 1955). In other words, an instrument is valid if an observed variance in the measurement is caused by an actual variation in the underlying theoretical construct (Heggestad et al., 2019, p. 2598). To establish construct validity, researchers examine convergent and discriminant validity. Construct validity is achieved when different measures for the same theoretical construct yield the same or similar results (i.e., convergent validity), while measures for other theoretical constructs yield different results (i.e., discriminant validity) (Braun & Kuljanin, 2015, p. 523).

Content validity describes how adequately an instrument captures the conceptual ideas behind a latent construct. It assesses whether the instrument captures all theoretical facets of the focal construct but not those of other constructs (Colquitt et al., 2019). Criterion or predictive validity describes how well a construct predicts or correlates with related constructs in the nomological network, as known and theorized according to the literature. Assessing predictive validity requires examining the focal and the related constructs at the same time (Schmitz & Storey, 2020). Demonstrating the validity of measurement instruments is crucial for substantiating inferences to theory and establishing validity is a nontrivial task. To account for this, research into instrument development offers best practices to ensure instrument validity.

2.3 Instrument development

Traditional instrument development originated in psychology. It focuses on developing survey scales to operationalize theoretical concepts. Information systems scholars have adopted instrument development with a "gradual accumulation and refinement of existing practices" (Burton-Jones & Lee, 2017, p. 451). In the following, I present a brief reproduction of the widely accepted practices (Churchill, 1979; Cortina et al., 2020; Hinkin, 1998; MacKenzie et al., 2011; Straub & Gefen, 2004). Although multiple practices to develop instruments exist, I focus on those practices that also apply to digital trace data and omit others that are specific to survey instruments[1] (Table 1).

Table 1. Summary of the instrument development process

Step 1:	Define theoretical constructs
Step 2:	Check for existing constructs
Step 3:	Develop initial instrument version
Step 4:	Pre-test instrument and revise
Step 5:	Perform full validation
Step 6:	Replicate instrument in further studies

The first step of instrument development is construct definition. Researchers must define the scope and goals of the research and identify the theoretical constructs of interest in the domain (Hinkin, 1998). Construct definitions can be developed by conducting literature or qualitative research (Churchill, 1979). MacKenzie et al. (2011) and Podsakoff et al. (2016) give

[1] The omitted practices include manipulation checks (Hauser et al., 2018), autocorrelation screenings (Gottfried et al., 2022), attention checks and survey response rates (Eysenbach, 2004; Fan & Yan, 2010), assessing sampling and representativeness of surveys (Kraemer & Pinsonneault, 1993), self-reporting biases (Podsakoff & Organ, 1986), and the common method bias (Podsakoff et al., 2003).

recommendations for defining theoretical constructs. They suggest defining the construct domain and identifying potential attributes of the construct, distinguishing necessary and sufficient attributes. After a preliminary construct definition is reached, researchers should revise it with a focus on stability and the boundary conditions to related constructs within the nomological network. Finally, the construct should be clearly and concisely documented.

The second step is to check for existing measurement instruments. In general, researchers should reuse or adapt existing instruments that are reliable and valid (Boynton & Greenhalgh, 2004). Since existing instruments might not work in other contexts or populations, estimating and reporting the validity and reliability statistics remains important (Compeau et al., 2022; Straub, 1989). Researchers must check for sufficient evidence that the existing instrument can be applied in a new context (Cortina et al., 2020). They must specify modifications and demonstrate the instrument's validity and reliability because already small changes can affect the psychometric characteristics of the measures (Heggestad et al., 2019, p. 2600). Detailed guidelines for instrument adaptation are available (Cortina et al., 2020; Heggestad et al., 2019; Newman et al., 2016; Pillet et al., 2023). Conversely, developing new instruments requires justification for why existing instruments are insufficient (Compeau et al., 2022). A new instrument should be empirically or theoretically superior. For example, digital trace data-based instruments may be finer-grained and more exhaustive than survey instruments (empirical superiority) or provide an innovative measurement for a theory previously non-operationalizable (theoretical superiority). Researchers should report the quality of a newly developed instrument and compare it to existing measurements (Compeau et al., 2022).

The third step is to build a new instrument if required. Researchers should think broadly, exhaustively covering the theoretical construct (Cortina et al., 2020). After a preliminary version of the instrument has been developed, seeking feedback, revising, and iterating with subject matter experts (e.g., domain experts) while attending to content validity is important. The instrument must correspond to the theoretical construct and be distinct from others (Colquitt et al., 2019). Quantitative approaches to assessing content validity are straightforward to implement (Anderson & Gerbing, 1991; Hinkin & Tracey, 1999; Lawshe, 1975). Guidelines are available for items-based measurement instruments (DeVellis, 2016; Hinkin, 1998; MacKenzie et al., 2011; Petter et al., 2007; Schmitz & Storey, 2020; Schriesheim et al., 1993).

The fourth step is pre-testing (pilot testing) the instrument. A data sample of the preliminary instrument is collected together with related measures from the nomological network (Hinkin, 1998). The related measures allow for estimating the focal instrument's relationships to existing constructs within the nomological network to assess predictive validity. Having collected the data, researchers conduct an exploratory factor analysis. The factor loadings and goodness of fit should be checked (Cortina et al., 2020; MacKenzie et al., 2011). Striving for parsimony, factors with low loading, low correlations, or low variance inflation factors should be removed. The reliability of the focal instrument is assessed by estimating the internal consistency of the measurement (Straub, 1989). Historically, Cronbach's alpha has been used (Cronbach & Meehl, 1955). However, McDonald's omega, also called congeneric reliability, or other reliability measures are recommended today (Cho, 2021; McNeish, 2018). Researchers should report and assess descriptive statistics. Based on the results, they can tweak the instrument by removing or modifying poor factors.

The fifth step is validation of the instrument. A second data sample is collected to conduct a confirmatory factor analysis to substantiate the instrument's robustness. Statistical tests exist to estimate convergent and discriminant validity (Straub, 1989; Straub & Gefen, 2004), for example, the multitrait-multimethod matrix (MTMM) (Campbell & Fiske, 1959), confirmatory factor analysis (Hair et al., 2019), or newer and recommended approaches such as the heterotrait-monotrait ratio of correlations (HTMT) (Henseler et al., 2015) or the updated HTMT2 (Roemer et al., 2021). Not only empirics but also theoretical reasoning can support the measures' construct validity (Cook & Campbell, 1979; Smith, 2005). Cortina et al. (2020) and MacKenzie et al. (2011) provide overviews of the statistical approaches for estimating convergent and discriminant validity. Beyond convergent and discriminant validity, researchers should demonstrate the predictive validity (or criterion-related validity) by estimating relationships between related constructs from the nomological network and the focal instrument (Compeau et al., 2022). Researchers should check the external validity of the instrument by estimating goodness of fit tests and comparing different models with variants of the focal instrument and the related constructs (Hinkin, 1998). Finally, researchers can assess the test-retest reliability, i.e., check whether the measurement is stable over time (Cortina et al., 2020). If the instrument is valid and reliable, it can be used in empirical research to draw inferences. Reuse of the developed instrument by other researchers will further substantiate its reliability and validity across generalized samples from other contexts and populations (Cortina et al., 2020; Hinkin, 1998; MacKenzie et al., 2011).

Despite the extensive literature on instrument development, it remains unclear whether these guidelines can be applied to digital trace data. Digital trace data constitute a different type of data compared to traditional survey scales; they do not consist of items designed by researchers that are meticulously developed and refined. Hence, it is important to consider the unique properties of digital trace data when developing digital trace data-based instruments.

3 Literature review

In this section, I describe how I reviewed the literature to identify potential issues and guidelines when using digital trace data for variance-based research. First, the current critiques are collected and analyzed to make sense of potential problems when using digital trace data as measurement instruments. For this, I synthesize the literature on digital trace data from the information systems and adjacent disciplines (e.g., organizational research, psychology). Second, I describe directions for future applications of digital trace data in hypothetico-deductive inquiries. To this end, I integrate the findings from the IS field with measurement theory from psychology.

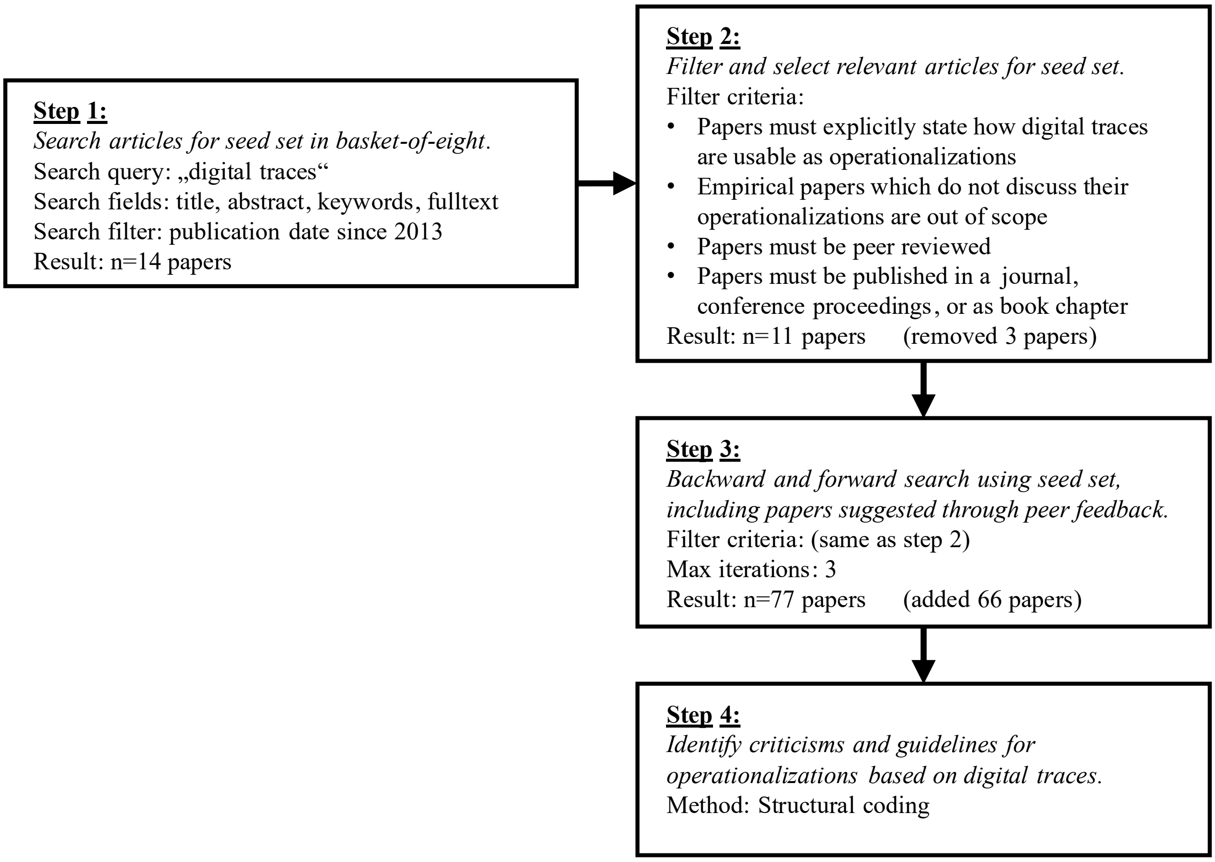

Figure 2. Research method

To conduct the literature review, I followed the guidelines by Schryen (2015). Reproducibility is established through the documentation of the review process, as captured in Figure 2. The literature review is representative and not exhaustive (Cooper, 1988). It focuses on backward and forward searches because the terminology surrounding digital trace data differs, and papers are found across adjacent disciplines (e.g., organizational research, psychology). I identified a seed set of papers from which backward and forward search iterations were conducted. To create the literature seed set, I searched the AIS senior scholar's basket of eight using the AIS eLibrary with the search query "digital traces," including papers since 2013. This search yielded n=11 relevant papers (n=14 total) (marked with (*) in the references). After the backward and forward search, I identified n=77 relevant papers. The two stopping criteria for the backward and forward search iterations were: (1) no more relevant papers, or (2) at max. 3 levels deep. This balanced the search effort with the marginal returns of analyzing more papers (vom Brocke et al., 2015).

I further filtered for peer-reviewed papers that explicitly state how digital trace data are used as measurement instruments to reach an overview of approaches to operationalizing theory using digital trace data. Table 3 and Table 4 describe the distribution of identified papers across disciplines and journals. Figure 3 illustrates the set of papers identified in the literature by the publication year. It is apparent that the number of articles based on digital trace data is increasing and that the analysis of digital trace data is solidifying itself as a common research approach.

Table 3. Distribution of surveyed papers across disciplines (based on VHB JourQual 3)

Discipline (based on VHB JourQual 3)	**Count**
Information Systems	47
Information Systems Conference Proceedings	10
Other	7
Management	4
Book Chapters	4
Natural Sciences	3
Psychology	2

Table 4. All cited outlets with more than 1 citation

Journal	**Count**
Journal of the Association for Information Systems	9
MIS Quarterly	7
Book chapters	5
Decision Support Systems	5
Information Systems Research	3
Journal of Strategic Information Systems	3
Proceedings of the ICIS	3
Information and Organization	3
European Journal of Information Systems	2
Organizational Research Methods	2
Business & Information Systems Engineering	2
Proceedings of the AMCIS	2
Science	2
Other	29

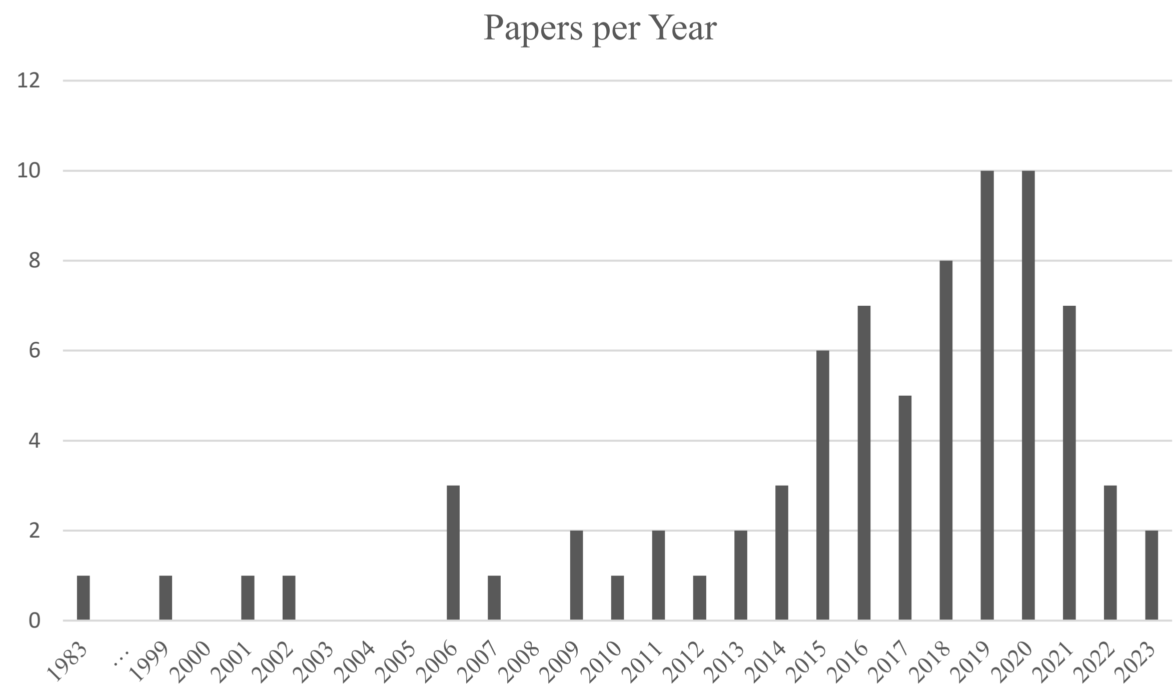

Figure 3. Papers on digital trace data per year

I applied structural coding to analyze the papers and identify the critique related to digital trace data-based construct measurements (Saldana, 2009). The segments with similar statements are merged and analyzed to form a coherent argument. Drawing from these arguments, I outline the central opportunities and critique of using digital trace data as measurement instruments.

4 Potential issues of using digital trace data for variance-theoretic research

Digital trace data occur at an unprecedented scale and granularity (Chaffin et al., 2017). Combined with novel computational analyses, they present ample opportunities for research on digital phenomena (Berente et al., 2018). Comparing digital trace data to self-reported data, Scharkow (2016) asserts that digital trace data may be more accurate. Perceptual responses, such as surveys or interviews, are subjective and sparse[2]. For example, surveys lack continuity

[2] Scharkow does not criticize rich qualitative methods per se. Researchers often seek subjective and perceptual data deliberately.

even if they are paneled (Eagle & Pentland, 2006). The changing nature of work towards distributed and virtual work, crossing spatial and temporal boundaries, hinders interviews and ethnographic accounts (Barley & Kunda, 2001). It is challenging to observe distributed individuals over extended periods. Digital trace data, therefore, promise a less skewed and more complete account of the historical behaviors of humans than surveys (Hüllmann, 2019).

Notwithstanding the opportunities, digital trace data should not be considered a panacea. While directly observable digital behaviors are a low-hanging fruit, e.g., measuring actual technology use over perceptions, measuring more abstract or latent constructs is difficult, e.g., attitudes or user perceptions (Howison et al., 2011). Despite the promise to capture arbitrary "behavioral constructs" (Chaffin et al., 2017; Hedman et al., 2013), more cautious voices express concerns regarding the establishment of proper procedures for analyzing digital trace data that ensure the validity of inferences (e.g., Grover et al., 2020; Johnson et al., 2019). In the following, I will address concerns related to the data-generating process, construct validity, the role of theory, as well as privacy and open access.

4.1 Data generation process

Grover et al. (2020) claim that digital trace data are often analyzed under an implied positivist paradigm. Despite a high number of observations, digital trace data remain reductionist and should be understood as signals or indicators, not truth (Freelon, 2014; Howison et al., 2011). They provide only a reflection of behaviors and are not objective (Østerlund et al., 2020). Instead, they can be interpreted in multiple—potentially opposing—ways (Freelon, 2014). As a result, the context and conditions under which digital trace data are generated must be considered for analysis (Flyverbom & Murray, 2018). Digital trace data are performative. The routine use of complex hardware and software systems generates digital trace data. When, how, and how much these systems are used depends on the organization in which they are deployed—the structures, context, and specific situation shape the human behavior and the subsequent trace generation (Aaltonen & Stelmaszak, 2023; Andersen et al., 2016).

The absence of digital trace data does not mean an absence of activity (Andersen et al., 2016). The system's capabilities define what actions and behaviors generate a trace. A sudden absence of digital trace data may occur due to shutting down a system, cleansing log files, faulty storage, or a system outage (Xu et al., 2020). Incomplete traces may bias the results and render them inconsistent, useless or misleading. Software and hardware systems are dynamic, and which behaviors are logged may change (Howison et al., 2011). Thus, the validation of digital trace data as a measurement instrument is an ongoing effort (Chaffin et al., 2017). Proprietary hardware and software systems exacerbate the problem of the measurement instrument's quality because the internal mechanisms are opaque, and researchers cannot check how the system works and generates digital trace data.

Beyond the systems that researchers can investigate, backchannel systems or shadow IT may generate data that are not analyzed (Hüllmann, 2022). To address this, researchers suggest looking for anomalies or sudden changes in the data (Howison et al., 2011), checking the quality and reliability of the hardware and software systems generating the data (Chaffin et al., 2017), and consulting subject matter experts with intimate knowledge of the hardware and software systems (Howison et al., 2011, p. 20).

Ensuring the quality and reliability of measurements depends on the specific device being deployed (Tonidandel et al., 2018). Different vendors, or different devices from the same vendor, yield different measurements. For example, fitness bracelets that measure your pulse frequency can vary, and the generated timestamps for different chat software can also vary. Hardware sensors suffer from noise, sensor sensitivity and configuration, or misapplications of the sensor (Chaffin et al., 2017). Ultimately, attributing variance in digital trace data that is caused by the measurement instrument to individuals and their behavior may render a study's result invalid (Chaffin et al., 2017).

Preprocessing digital trace data includes decisions on cleansing, normalizing, or transforming the data, e.g., removing all outliers beyond a certain threshold. But such thresholds, or preprocessing decisions, are often arbitrary and not theoretically guided (Xu et al., 2020). Researchers should either argue for their preprocessing decisions, grounded in theoretical assumptions, or report how alternative decisions would have affected the results in terms of a sensitivity analysis.

Analyzing large sets of digital trace data comes with peculiarities that require caution. Typically, the studies are overpowered due to a high number of observations, resulting in conflated significance values (e.g., p-values) (Johnson et al., 2019). Hence, attention should be paid to the estimated effect sizes and the data-generating process (Mertens & Recker, 2020). For dyadic trace data, the assumptions of parametrical statistical analysis are often ignored and violated (Howison et al., 2011; Hüllmann & Kroll, 2018). For example, social interactions on an enterprise social network or chat platform are by definition not identically and independently distributed, for if one sends a message, another one will receive a message.

4.2 Construct validity

Digital trace data are logs of historical human actions and behaviors that are not necessarily generated for research purposes. To derive theoretical insights from digital trace data, researchers must link the measurement instruments (i.e., digital trace data) to higher-level theoretical constructs (Chaffin et al., 2017; Howison et al., 2011). Establishing the link between digital trace data and theoretical constructs is not straightforward (Chaffin et al., 2017, p. 6). The operationalization of the theoretical constructs is based on human design choices and must be argued and tested (Chaffin et al., 2017; Lindberg, 2020). However, an argument for the operationalization is often missing, and authors seldom address the construct validity in digital trace data research (Braun & Kuljanin, 2015; Howison et al., 2011). For other instruments, such as surveys, construct validity is meticulously established through tests. For digital trace data, establishing construct validity solely through quantitative instruments and statistical means is difficult because it is unclear what the baseline or ground truth is (Braun & Kuljanin, 2015).

4.3 The role of theory

Studies relying on digital trace data have suffered from the *streetlight effect*, that is, they have favored research problems for which data are readily available over those that require substantive research (Rai, 2017). These studies avoid latent constructs and favor behavioral constructs that are easily measurable at scale instead (Johnson et al., 2019). For example, *technology use* is directly measurable and would be favored over *perceived usefulness of technology* which is not directly measurable. The studies try to impress with big data sets instead of addressing meaningful research problems that add to cumulative knowledge (Grover et al., 2020). The original theory is of secondary importance and chosen ad-hoc or post-hoc to fit the patterns in the data. Howison et al. (2011) call this *theoretical fitting*. Landers et al. (2016, p. 480) put it pejoratively and insist on avoiding *brute force empiricism* with digital trace data research. Studies without robust theorizing tend to be incremental and narrowly empirical (Grover et al., 2020, p. 277) and do not generalize well from the specific context in which the data was collected (Johnson et al., 2019). They are merely addressing local problems and obstruct the purpose of building generalizable knowledge, preventing meaningful theoretical contributions.

4.4 Privacy and open access

Privacy concerns depend on the mode of digital trace data generation. The *active* and deliberate generation of digital trace data for research purposes can include informed consent from the study participants *a priori*. Participants can voluntarily share their active traces from only a limited observation period, e.g., they wear a sociometric badge for two weeks. *Passive* digital trace data, on the contrary, are historical logs generated as a by-product of using a software or device. They potentially span long observation periods and may not be deleted or cleared for years. Researchers can only ask for consent *a posteriori* after the data has been generated. When researchers use passive traces, individuals may not know that their data are being used for research, which poses ethical concerns (Markus & Marabelli, 2017; Tonidandel et al., 2018). For example, studies conducted on social media platforms such as X.com (Twitter) or Yammer typically do not inform users that their data are used for research purposes.

Long-term and detailed accounts of human behavior are sensitive, and workplace data are proprietary. Often, the data cannot be shared open-access, preventing independent researchers from replicating the results (Tonidandel et al., 2018).

5 Guidelines for developing measurements with digital trace data

The discussed concerns highlight that guidelines for operationalizing theoretical constructs using digital trace data are needed, especially in light of the surging empirical applications vis-à-vis the IS field lacking established procedures. The quality of measurement instruments significantly affects the inferential results, and robust measurement theory can ensure valid and reliable construct operationalizations. The procedure of designing digital trace data-based instruments differs from traditional ones, such as survey scales. The argument of digital trace data's superiority over traditional instruments fails to recognize that surveys and other established instruments have a history of being reliable, valid, and robust across contexts and populations, as replications have shown (Cortina et al., 2017, 2020). The extensive measurement literature, which deems it essential to provide quantitative evidence about instrument validity and reliability, can inspire digital trace data research (Cortina et al., 2020). Drawing from this literature, I craft recommendations for developing digital trace data-based instruments. The recommendations are an amalgamation of the identified concerns with digital trace data research and how best practices from measurement theory can address them (Table 5). The recommendations are structured along six steps that offer the IS field guidance to validate digital trace data-based instruments.

Table 5. Validation guidelines for digital trace data

<u>Step 1 – Define theoretical construct:</u> • Define theoretical constructs • Avoid being purely data-driven • Keep in mind which kind of digital trace data exist in the empirical setting • Use literature work and subject matter experts (qualitative) **<u>Step 2 – Check for existing instruments:</u>** • Check for existing instruments (both digital trace data-based and traditional) • Check if existing instruments apply to the empirical setting • Check if digital trace data-based instruments can be empirically or theoretically superior • Adapt instrument if applicable • Follow guidelines for adapting instruments (Heggestad et al., 2019; Pillet et al., 2023) • Use literature work and subject matter experts (qualitative) **<u>Step 3 – Develop preliminary instrument:</u>** • Develop initial instrument version • Check for face validity and content validity • Scrutinize the ETL process and gauge data quality • Explain the nature of digital trace data and their correspondence to the theoretical construct • Use subject matter experts (qualitative) • Collect a first sample (include instruments for other constructs and using other methods) • Compute exploratory factor analysis, MTMM, or alternative tests • Estimate content validity via subject matter experts (quantitative) • Tweak the ETL process and revise the instrument as needed **<u>Step 4 – Validate instrument:</u>** • Perform full validation • Collect a second sample (include instruments for other constructs and using other methods) • Estimate construct validity (convergent and discriminant validity) via confirmatory factors analysis, MTMM, or alternative tests • Estimate content validity via subject matter experts (quantitative) • Estimate predictive validity (criterion validity) by estimating path coefficients to theoretically related constructs • Estimate goodness of fit for external validity **<u>Step 5 – Report instrument:</u>**

- Report details on the ETL process and instrument validation procedure
- Report software used and test statistics
- For ETL process report (Marsden & Pingry, 2018; Vial, 2019):
    - How and when data was extracted and transformed
    - Source code for ETL process
    - Provide copies of the original dataset
    - If applicable: API descriptions and potential restrictions

Step 6 – Replicate instrument:

- Replication: other researchers adopting the instrument will enhance generalizing its reliability and validity across contexts and populations (Hinkin, 1998)

Step 1 – Define theoretical construct: First, digital trace data research should not be conducted agnostic to theory. Although data-driven approaches are popular, engaging with theory remains critical. In a deductive variance-theoretic setting, engaging with the literature to concisely define and demarcate theoretical constructs is essential (MacKenzie et al., 2011). Without theoretical clarity, developing robust instruments based on digital trace data is impossible (Suddaby, 2010). Instead of being exclusively data-driven, the thorough analysis of the literature and triangulation with qualitative data help avoid the *streetlight effect* and merely engaging in a computational exercise (Rai, 2017; Thapa et al., 2021).

The empirical setting of the study should be linked to a broader knowledge goal that informs what digital trace data can lead to a theoretical contribution (Johnson et al., 2019). For studies with digital trace data, it should be considered that theory from traditional studies in the offline context does not necessarily generalize to a digital context, as long as it is not evidenced through empirical data (Hüllmann, 2022). For longitudinal studies, researchers should address whether the phenomenon under study is stable for the observation period so that a variance theory based on digital trace data can explain it (Howison et al., 2011). Only looking for patterns in data does not necessarily produce generalizable insights that contribute to broader knowledge. On the contrary, sound theoretical constructs are the first step to facilitating contributions by revising and improving theory through digital trace data.

Step 2 – Check for existing instruments: After the theoretical constructs have been defined, researchers can develop a measurement instrument based on digital trace data. First, they should check for existing instruments. If a digital trace data-based instrument exists, the validation and reliability may already have been established, and reusing the instrument will contribute to substantiating its quality (Heggestad et al., 2019; Pillet et al., 2023). However, the literature review showed that this practice is seldomly followed for digital trace data. Compared to traditional instruments, reusing digital trace data instruments is more challenging due to data-generating processes that can be unique to a sampled organization and the configurations of its systems.

Since digital trace data research rarely reports validity and reliability, it is difficult to assess an existing instrument's quality (Braun & Kuljanin, 2015). Thus, when adopting an established digital trace data instrument, reassessing its validity and reliability remains essential. Instrument validity does not necessarily translate from one context to another, as configurations of information systems vary. Another instrument may exist based on traditional approaches, for example, surveys. Researchers should only prefer the digital trace data-based instrument if validity and reliability tests show that it is empirically or theoretically superior to existing instruments (Pillet et al., 2023). Furthermore, I encourage triangulation with subject matter experts. Subject matter experts have intimate insights into the hardware and software systems that generate digital trace data. They can be the developers, designers, owners, or power users of the system with technical knowledge of what actions generate what kind of digital trace data. Conversing with multiple subject matter experts specializing in different subsystems may be necessary for very complex systems.

Step 3 – Develop preliminary instrument: When a new measurement needs to be designed, researchers must identify suitable types of digital trace data instead of generating items. Triangulating literature reviews with the consultation of subject matter experts can be a key lever for coming up with suitable digital trace data types and the specific event logs that measure the theoretical construct in question. The literature on data quality and information systems engineering helps to identify such digital trace data (Vial, 2019). Although digital trace data are hailed as more objective, they remain reductionist and only indicators of behaviors and latent constructs (Freelon, 2014). Understanding the data-generating process is imperative for testing face and content validity. Intimate knowledge about the system is needed to assess the nature of the digital trace data and gauge data quality. *An absence of digital trace data is not an absence of behavior.* Only subject matter experts and the owners of the information systems can provide the necessary details on proprietary systems, potential system outages, and what behaviors generate what kind of digital trace data (Flyverbom & Murray, 2018; Xu et al., 2020).

I conjecture that engaging with subject matter experts is more critical in developing digital trace data instruments than surveys because of their intimate knowledge of how the source information system(s) work and are being used. Consequently, face and content validity are important in digital trace data research because digital trace data as *passive exhaust* may be generated through various actions (Aaltonen & Stelmaszak, 2023). For passive digital trace data, the data-generating process is not controllable, contrary to active digital trace data, where researchers can influence the data generation. Similarly to the identification of appropriate digital trace data, researchers can demonstrate content validity through subject matter experts and data quality assessments (Marsden et al., 2019; Vial, 2019).

The development and application of traditional measurements are highly paradigmatic (Cortina et al., 2020). Survey-based papers describe the context of data collection, the sample and population, and report the scales and methods used. They add analysis details where necessary, such as data transformations (e.g., log transformation), removal of outliers, or testing parametric assumptions (Aguinis et al., 2018). Conversely, the development and application of digital trace data-based instruments are non-paradigmatic due to unique *extract, transform, and load* (ETL) processes. The ETL process severely impacts the measurement instrument's quality at design time before empirical application. Thus, for designing digital trace data-based instruments, it is imperative to report how the data was extracted, transformed, loaded, and preprocessed to yield the final numerical measurements (Pipino et al., 2002; Vial, 2019). Common statistical assumptions for regression-based analysis may not hold for digital trace data instruments (Howison et al., 2011). For example, dyadic digital trace data are not identically and independently distributed. Such issues should be reported when designing a digital trace data-based measurements for variance-theoretic research because it affects the inferential procedures.

Step 4 – Validate instrument: Assessing the ETL process and data quality is necessary but insufficient for validating a digital trace data instrument. Statistical testing of instrument validity and reliability remains critical. Not all, but selected approaches from the measurement literature can be adopted for digital trace data until dedicated methods are available. Most of these methods require composite measures, but digital trace data are often used as single-item measures. Thus, I draw from the literature on validating single-item measures for crafting the digital trace data guidelines (Allen et al., 2022; Fuchs & Diamantopoulos, 2009).

Content validity can be estimated using established approaches as recommended (Anderson & Gerbing, 1991; Colquitt et al., 2019; Lawshe, 1975). Correlations and the multitrait-multimethod matrix (MTMM) can be applied to establish convergent and discriminant validity for digital trace data instruments (Campbell & Fiske, 1959; Matthews et al., 2022). For MTMM, the single-item digital trace data construct can be compared to an existing and already validated multi-item measure. This approach works if at least one other multi-item measure is available (Fuchs & Diamantopoulos, 2009; Matthews et al., 2022). Confirmatory factor analysis can be used by including both the digital trace data single-item and the multi-item construct and having both load on the same latent factor. Convergent validity cannot be assessed if no multi-item constructs for the same theoretical construct are available.

Single-item measures can be regressed on related constructs for estimating the predictive validity (or criterion validity) (Allen et al., 2022; Fuchs & Diamantopoulos, 2009; Matthews et al., 2022). Beyond test-retest reliability, the temporal stability of digital trace data can be examined (DeVellis, 2016). Other validity testing approaches are currently being developed. For example, Matthews et al. (2022) suggest validating a highly correlative proxy instrument via multi-item scales. Researchers can use common guidelines for reporting the statistical tests (MacKenzie et al., 2011). Despite the availability of statistical tests, established cutoff points remain unknown for digital trace data (Lance et al., 2006; Matthews et al., 2022). Until such cutoff points are established, qualitative arguments remain key next to the reporting of statistics. Triangulation with other instruments and subject matter experts can corroborate the measurements' robustness.

Step 5 – Report instrument: Given the unique nature of digital trace data, it is important to detail the ETL process and instrument validation procedure extensively. Report how and when the data was extracted and how it was transformed, specifying the source information systems, the used tools, and the necessary context. If possible, provide the source code for the ETL process and copies of the original and transformed data sets. If applicable, provide the necessary API descriptions and potential restrictions that may apply for researchers looking to reproduce the data collection. Finally, if numerical measures have been derived from the digital trace data, report all customary summary and test statistics as outlined in the best practices of measurement theory.

Step 6 – Replicate instrument: Publishing your instrument with detailed insights about its development and validation will offer other researchers to adapt your instrument in subsequent studies, contributing evidence toward the instrument's validity. Reusing the instrument in other contexts and applying it to other populations will enhance its generalizability and reliability.

6 Limitations

Although measurement theory provides the inspiration for this chapter, it deals primarily with survey-based items and, therefore, is limited as a frame of reference for building instruments based on digital trace data (Cortina et al., 2020). In particular, single-item validation has been labeled “a process as much art as science” (Matthews et al., 2022, p. 669). Albeit single-item measures have been shown to be as valid and reliable as multi-item measures (Allen et al., 2022), dedicated research is needed to conceive statistical approaches for digital trace data-based instrument validation. The recommendations (Table 5) focus on the instrument validity after numerical measures have been computed. The ETL process, however, is critical for digital trace data-based instruments. Researchers should examine the analysis and preprocessing configurations and the context of the data-generating process to see if contextual factors, e.g., system failure, internet outages, or other reasons, bias the results. Future research can investigate how to mitigate these procedural and contextual causes by triangulation with qualitative approaches.

Even meticulously reporting the ETL process does not guarantee the digital trace data instruments’ reliability and subsequent reusability. Contrary to traditional instruments, digital trace data instruments are hard to reuse unless the data set is curated and published for others. The reason is that researchers extract digital trace data from a variety of complex hardware and software systems with unique configurations and data-generating processes. Although researchers exert control over *active* digital trace data, the level of access to *passive* digital trace data varies. Public APIs can be restricted, e.g., the Meta or X.com (Twitter) APIs, and the availability of proprietary company data depends on organizational clearance. Detailed insights into an ETL process might have to remain confidential if they exhibit sensitive information about critical company operations. Furthermore, digital trace data instrument development comes with significant researcher degrees of freedom, which runs against the conventional wisdom of how we view the highly paradigmatic development of traditional instruments, e.g., surveys. Future research can investigate how digital trace data research, especially the ETL process, can be streamlined and standardized to foster a cumulative tradition (e.g., Grisold et al., 2023).

7 Conclusion and future work

Robust measurement instruments are crucial for theory testing (Dennis & Valacich, 2014). While previous research illuminated how to analyze digital trace data (Miranda et al., 2022; Pentland et al., 2021), the IS field has provided little guidance to ensure that findings are robust, valid, and reliable in variance-theoretic settings. Drawing from measurement theory, this chapter has provided guidelines for assessing and reporting the validity and reliability of digital trace data-based instruments for variance-theoretical empirical research. The guidelines comprise six steps that aid researchers from coming up with theoretically meaningful concepts to developing and reporting robust instruments based on digital trace data. Common concerns in digital trace data research such as construct validity, the role of theory, and the data-generating process are covered. Researchers can use the guidelines for assessing the instrument validity of digital trace data and avoiding spurious findings and replication issues, while contributing to cumulative knowledge.

8 Acknowledgments

My sincere thanks go to Bastian Wurm, who has repeatedly invested significant efforts to improve the chapter. Further valuable feedback was received by Joseph Nserat, Jan Mendling, and the two anonymous reviewers of the chapter. Finally, I appreciate the opportunity to contribute to this important book.

References

Aaltonen, A., & Stelmaszak, M. (2023). The Performative Production of Trace Data in Knowledge Work. *Information Systems Research*, *September*. https://doi.org/10.1287/isre.2019.0357

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What You See Is What You Get?

Enhancing Methodological Transparency in Management Research. *Academy of Management Annals*, *12*(1), 83–110. https://doi.org/10.5465/annals.2016.0011

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single Item Measures in Psychological Science. *European Journal of Psychological Assessment*, *38*(1), 1–5. https://doi.org/10.1027/1015-5759/a000699

Andersen, J. V., Lindberg, A., Lindgren, R., & Selander, L. (2016). Algorithmic Agency in Information Systems: Research Opportunities for Data Analytics of Digital Traces. *Proceedings of 49th Hawaii International Conference on System Sciences (HICSS)*, 4597–4605. https://doi.org/10.1109/HICSS.2016.571

Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, *76*(5), 732–740. https://doi.org/10.1037/0021-9010.76.5.732

(*) Avital, M., Jensen, T. B., & Dyrby, S. (2021). The social fabric framework: steps to eliciting the social making of organisations in the digital age. *European Journal of Information Systems*, 1–27. https://doi.org/10.1080/0960085X.2021.1907236

(*) Badakhshan, P., Wurm, B., Grisold, T., Geyer-Klingeberg, J., Mendling, J., & vom Brocke, J. (2022). Creating business value with process mining. *The Journal of Strategic Information Systems, 31*(4), 101745. https://doi.org/10.1016/j.jsis.2022.101745

Barley, S. R., & Kunda, G. (2001). Bringing Work Back In. *Organization Science*, *12*(1), 76–95. https://doi.org/10.1287/orsc.12.1.76.10122

Berente, N., Seidel, S., & Safadi, H. (2018). Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research*, *30*(1), 50–64. https://doi.org/10.1287/isre.2018.0774

Boynton, P. M., & Greenhalgh, T. (2004). Selecting, designing, and developing your questionnaire. *BMJ*, *328*(7451), 1312–1315. https://doi.org/10.1136/bmj.328.7451.1312

Braun, M. T., & Kuljanin, G. (2015). Big Data and the Challenge of Construct Validity. *Industrial and Organizational Psychology*, *8*(4), 521–527. https://doi.org/10.1017/iop.2015.77

Burton-Jones, A., & Lee, A. S. (2017). Thinking About Measures and Measurement in Positivist Research: A Proposal for Refocusing on Fundamentals. *Information Systems Research*, *28*(3), 451–467. https://doi.org/10.1287/isre.2017.0704

Burton-Jones, A., McLean, E. R., & Monod, E. (2015). Theoretical perspectives in IS research: from variance and process to conceptual latitude and conceptual fit. *European Journal of Information Systems*, *24*(6), 664–679. https://doi.org/10.1057/ejis.2014.31

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Chaffin, D., Heidl, R., Hollenbeck, J. R., Howe, M., Yu, A., Voorhees, C., & Calantone, R. (2017). The Promise and Perils of Wearable Sensors in Organizational Research. *Organizational Research Methods*, *20*(1), 3–31. https://doi.org/10.1177/1094428115617004

Cho, E. (2021). Neither Cronbach's Alpha nor McDonald's Omega: A Commentary on Sijtsma and Pfadt. *Psychometrika*, *86*(4), 877–886. https://doi.org/10.1007/s11336-021-09801-1

Churchill, G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, *16*(1), 64–73. http://www.jstor.org/stable/3150876

Colquitt, J. A., Sabey, T. B., Rodell, J. B., & Hill, E. T. (2019). Content validation guidelines: Evaluation criteria for definitional correspondence and definitional distinctiveness. *Journal of Applied Psychology*, *104*(10), 1243–1265. https://doi.org/10.1037/apl0000406

Compeau, D., Correia, J., & Thatcher, J. (2022). When Constructs Become Obsolete: A Systematic Approach to Evaluating and Updating Constructs for Information Systems Research. *MIS Quarterly*, *46*(2), 679–712. https://doi.org/10.25300/MISQ/2022/15516

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (1st ed.). Houghton Mifflin.

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, *1*(1), 104–126. https://doi.org/10.1007/BF03177550

Cortina, J. M., Aguinis, H., & DeShon, R. P. (2017). Twilight of dawn or of evening? A century of research methods in the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*(3), 274–290. https://doi.org/10.1037/apl0000163

Cortina, J. M., Sheng, Z., Keener, S. K., Keeler, K. R., Grubb, L. K., Schmitt, N., Tonidandel, S., Summerville, K. M., Heggestad, E. D., & Banks, G. C. (2020). From alpha to omega and beyond! A look at the past, present, and (possible) future of psychometric soundness in the Journal of Applied Psychology. *Journal of Applied Psychology*, *105*(12), 1351–1381. https://doi.org/10.1037/apl0000815

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

Dennis, A., & Valacich, J. (2014). A Replication Manifesto. *AIS Transactions on Replication Research*, *1*(August 2014), 1–4. https://doi.org/10.17705/1atrr.00001

DeVellis, R. F. (2016). *Scale Development: Theory and Applications* (L. Bickman & D. J. Rog (eds.); 4th ed.). SAGE Publications Ltd.

Eagle, N., & Pentland, A. (Sandy). (2006). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, *10*(4), 255–268. https://doi.org/10.1007/s00779-005-0046-3

Eysenbach, G. (2004). Improving the Quality of Web Surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research*, *6*(3), e34. https://doi.org/10.2196/jmir.6.3.e34

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, *26*(2), 132–139. https://doi.org/10.1016/j.chb.2009.10.015

Flake, J. K. (2021). Strengthening the foundation of educational psychology by integrating construct validation into open science reform. *Educational Psychologist*, *56*(2), 132–141. https://doi.org/10.1080/00461520.2021.1898962

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Flyverbom, M., & Murray, J. (2018). Datastructuring—Organizing and curating digital traces into action. *Big Data & Society*, *5*(2), 1–12. https://doi.org/10.1177/2053951718799114

Freelon, D. (2014). On the Interpretation of Digital Trace Data in Communication and Social Computing Research. *Journal of Broadcasting & Electronic Media*, *58*(1), 59–75. https://doi.org/10.1080/08838151.2013.875018

Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft*, *69*(2), 195–211.

Gottfried, J., Ježek, S., Králová, M., & Řiháček, T. (2022). Autocorrelation Screening: A Potentially Efficient Method for Detecting Repetitive Response Patterns in Questionnaire Data. *Practical Assessment, Research and Evaluation*, *27*(1). https://doi.org/10.7275/vyxb-gt24

(*) Grisold, T., Kremser, W., Mendling, J., Recker, J., Brocke, J. vom, & Wurm, B. (2023). Keeping pace with the digital age: Envisioning information systems research as a platform. *Journal of Information Technology*, *38*(1), 60–66. https://doi.org/10.1177/02683962221130429

(*) Grover, V., Lindberg, A., Benbasat, I., & Lyytinen, K. (2020). The Perils and Promises of Big Data Research in Information Systems. *Journal of the Association for Information Systems*, *21*(2), 268–293. https://doi.org/10.17705/1jais.00601

(*) Hafermalz, E., Johnston, R. B., Hovorka, D. S., & Riemer, K. (2020). Beyond 'mobility' A new understanding of moving with technology. *Information Systems Journal, 30*(4), 762–786. https://doi.org/10.1111/isj.12283

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage.

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are Manipulation Checks Necessary? *Frontiers in Psychology*, *9*(JUN), 1–10. https://doi.org/10.3389/fpsyg.2018.00998

Hedman, J., Srinivasan, N., & Lindgren, R. (2013). Digital Traces of Information Systems: Sociomateriality made Researchable. *Proceedings of the 34th International Conference on Information Systems (ICIS)*, 1–17.

Heggestad, E. D., Scheaf, D. J., Banks, G. C., Monroe Hausfeld, M., Tonidandel, S., & Williams, E. B. (2019). Scale Adaptation in Organizational Science Research: A Review and Best-Practice Recommendations. *Journal of Management*, *45*(6), 2596–2627. https://doi.org/10.1177/0149206319850280

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods*, *1*(1), 104–121. https://doi.org/10.1177/109442819800100106

Hinkin, T. R., & Tracey, J. B. (1999). An Analysis of Variance Approach to Content Validation. *Organizational Research Methods*, *2*(2), 175–186. https://doi.org/10.1177/109442819922004

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity Issues in the Use of Social Network Analysis for the Study of Online Communities. *Journal of the Association for Information Systems*, *12*(12), 767–797. https://doi.org/10.1.1.157.9237

Hüllmann, J. A. (2019). The Construction of Meaning through Digital Traces. *Proceedings of* *the Pre-ICIS 2019, International Workshop on The Changing Nature of Work*.

Hüllmann, J. A. (2022). Media Choice in the Digital Era: A Replication Study using Digital Traces. *Proceedings of the 35th Bled EConference*.

Hüllmann, J. A., & Krebber, S. (2020). Identifying Temporal Rhythms using Email Traces. *Proceedings of the America's Conference of Information Systems (AMCIS)*.

Hüllmann, J. A., & Kroll, T. (2018). The Impact of User Behaviours on the Socialisation Process in Enterprise Social Networks. *Proceedings of the Australasian Conference on Information Systems (ACIS)*.

Johnson, S. L., Gray, P., & Sarker, S. (2019). Revisiting IS research practice in the era of big data. *Information and Organization*, *29*(1), 41–56. https://doi.org/10.1016/j.infoandorg.2019.01.001

Kraemer, K. L., & Pinsonneault, A. (1993). Survey research methodology in management information systems: An assessment. *Journal of Management Information Systems*, *10*(2), 75–105.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The Sources of Four Commonly Reported Cutoff Criteria. *Organizational Research Methods*, *9*(2), 202–220. https://doi.org/10.1177/1094428105284919

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological Methods*, *21*(4), 475–492. https://doi.org/10.1037/met0000081

Langley, A. (1999). Strategies for Theorizing from Process Data. *The Academy of Management Review*, *24*(4), 691–710. http://www.jstor.org/stable/259349

Lawshe, C. H. (1975). A Quantitative Approach To Content Validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

(*) Lindberg, A. (2020). Developing Theory Through Integrating Human and Machine Pattern Recognition. *Journal of the Association for Information Systems*, *21*(1), 90–116. https://doi.org/10.17705/1jais.00593

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques. *MIS Quarterly*, *35*(2), 293. https://doi.org/10.2307/23044045

Markus, M. L., & Marabelli, M. (2017). Researching Big Data Research: Ethical Implications for IS Scholars. *The 23rd Americas Conference on Information Systems (AMCIS)*, 1–5.

Marsden, J. R., & Pingry, D. E. (2018). Numerical data quality in IS research and the implications for replication. *Decision Support Systems*, *115*(October), A1–A7. https://doi.org/10.1016/j.dss.2018.10.007

Marsden, J. R., Pingry, D. E., & Thatcher, J. B. (2019). Perspectives on numerical data quality in IS research. *Decision Support Systems*, *126*(October), 113172. https://doi.org/10.1016/j.dss.2019.113172

Matthews, R. A., Pineault, L., & Hong, Y.-H. (2022). Normalizing the Use of Single-Item Measures: Validation of the Single-Item Compendium for Organizational Psychology. *Journal of Business and Psychology*, *37*(4), 639–673. https://doi.org/10.1007/s10869-022-09813-3

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Mertens, W., & Recker, J. (2020). New Guidelines for Null Hypothesis Significance Testing in Hypothetico-Deductive IS Research. *Journal of the Association for Information Systems*, *21*(4), 1072–1102. https://doi.org/10.17705/1jais.00629

Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences From Persons' Responses and Performances as Scientific Inquiry Into Score Meaning. *American Psychologist*, *50*(9), 741–749.

Miranda, S., Berente, N., Seidel, S., Safadi, H., & Burton-Jones, A. (2022). Editor's Comments: Computationally Intensive Theory Construction: A Primer for Authors and Reviewers. *MIS Quarterly*, *46*(2), iii–xviii.

(*) Müller, O., Junglas, I., Brocke, J. vom, & Debortoli, S. (2016). Utilizing big data analytics for information systems research: challenges, promises and guidelines. *European Journal of Information Systems, 25*(4), 289–302. https://doi.org/10.1057/ejis.2016.2

Narayanan, A. (2019). How to recognize AI snake oil. Presentation. In *Princeton Lectures* (pp. 1–21).

(*) Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification.' *The Journal of Strategic Information Systems*, *24*(1), 3–14. https://doi.org/10.1016/j.jsis.2015.02.001

Newman, D. A., Harrison, D. A., Carpenter, N. C., & Rariden, S. M. (2016). Construct Mixology: Forming New Management Constructs by Combining Old Ones. *The Academy of Management Annals*, *10*(1), 943–995. https://doi.org/10.1080/19416520.2016.1161965

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, *2*(13). https://doi.org/10.3389/fdata.2019.00013

Østerlund, C., Crowston, K., & Jackson, C. (2020). Building an Apparatus: Refractive, Reflective, and Diffractive Readings of Trace Data. *Journal of the Association for Information Systems*, *21*(1), 1–22. https://doi.org/10.17705/1jais.00590

Pentland, A. (2015). *Social Physics* (1st ed.). Penguin Books.

(*) Pentland, B. T., Recker, J., Wolf, J. R., & Wyner, G. (2020). Bringing context inside process research with digital trace data. *Journal of the Association for Information Systems*, *21*(5), 1214–1236. https://doi.org/10.17705/1jais.00635

(*) Pentland, B. T., Vaast, E., & Wolf, J. R. (2021). Theorizing Process Dynamics with Directed Graphs: A Diachronic Analysis of Digital Trace Data. *MIS Quarterly*, *45*(2), 967–984. https://doi.org/10.25300/MISQ/2021/15360

Petter, S., Straub, D. W., & Rai, A. (2007). Specifying Formative Constructs in Information Systems Research. *MIS Quarterly*, *31*(4), 623–656.

Pillet, J. C., Carillo, K. D., Vitari, C., & Pigni, F. (2023). Improving scale adaptation practices in information systems research: Development and validation of a cognitive validity assessment method. *Information Systems Journal*, *33*(4), 842–889. https://doi.org/10.1111/isj.12428

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211–218. https://doi.org/10.1145/505248.506010

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for Creating Better Concept Definitions in the Organizational, Behavioral, and Social Sciences. *Organizational Research Methods*, *19*(2), 159–203. https://doi.org/10.1177/1094428115624965

Podsakoff, P. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management*, *12*(4), 531–544. https://doi.org/10.1177/014920638601200408

Rai, A. (2017). Avoiding Type III Errors: Formulating IS Research Problems that Matter. *MIS Quarterly*, *41*(2), iii–vii.

Recker, J. (2021). *Scientific Research in Information Systems* (2nd ed.). Springer International Publishing. https://doi.org/10.1007/978-3-030-85436-2

Roemer, E., Schuberth, F., & Henseler, J. (2021). HTMT2–an improved criterion for assessing discriminant validity in structural equation modeling. *Industrial Management & Data Systems*, *121*(12), 2637–2650. https://doi.org/10.1108/IMDS-02-2021-0082

Rothmeier, K., Pflanzl, N., Hüllmann, J. A., & Preuss, M. (2021). Prediction of Player Churn and Disengagement Based on User Activity Data of a Freemium Online Strategy Game. *IEEE Transactions on Games*, *13*(1), 78–88. https://doi.org/10.1109/TG.2020.2992282

Saldana, J. (2009). *The coding manual for qualitative reseachers* (1st ed.). SAGE Publications Ltd.

Scharkow, M. (2016). The Accuracy of Self-Reported Internet Use—A Validation Study Using Client Log Data. *Communication Methods and Measures*, *10*(1), 13–27. https://doi.org/10.1080/19312458.2015.1118446

Schmitz, K., & Storey, V. C. (2020). Empirical Test Guidelines for Content Validity: Wash, Rinse, and Repeat until Clean. *Communications of the Association for Information Systems*, *47*(1), 787–850. https://doi.org/10.17705/1CAIS.04736

Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., & Lankau, M. J. (1993). Improving Construct Measurement In Management Research: Comments and a Quantitative Approach for Assessing the Theoretical Content Adequacy of Paper-and-Pencil Survey-Type Instruments. *Journal of Management*, *19*(2), 385–417. https://doi.org/10.1177/014920639301900208

Schryen, G. (2015). Writing Qualitative IS Literature Reviews—Guidelines for Synthesis, Interpretation, and Guidance of Research. *Communications of the Association for Information Systems*, *37*(August), 286–325. https://doi.org/10.17705/1CAIS.03712

Smith, G. T. (2005). On Construct Validity: Issues of Method and Measurement. *Psychological Assessment*, *17*(4), 396–408. https://doi.org/10.1037/1040-3590.17.4.396

Song, C., Qu, Z., Blumm, N., & Barabasi, A.-L. (2010). Limits of Predictability in Human Mobility. *Science*, *327*(5968), 1018–1021. https://doi.org/10.1126/science.1177170

Stewart, I. (2019). Our behaviour in bulk is more predictable than we like to imagine. *Aeon Essays*, 1–6.

Straub, D. W. (1989). Validating Instruments in MIS Research. *MIS Quarterly*, *13*(2), 147.

https://doi.org/10.2307/248922

Straub, D. W., & Gefen, D. (2004). Validation Guidelines for IS Positivist Research. *Communications of the Association for Information Systems*, *13*(March). https://doi.org/10.17705/1CAIS.01324

Suddaby, R. (2010). Editor's Comments: Construct Clarity In Theories Of Management And Organization. *Academy of Management Review*, *35*(3), 346–357. https://doi.org/10.5465/AMR.2010.51141319

Thapa, R., Sharma, P., Hüllmann, J. A., & Savarimuthu, B. T. R. (2021). Identifying Influence Mechanisms in Permissionless Blockchain Communities: The Bitcoin Case. *Proceedings of the 42nd International Conference on Information Systems (ICIS)*, 1–16.

Tonidandel, S., King, E. B., & Cortina, J. M. (2018). Big Data Methods: Leveraging Modern Data Analytic Techniques to Build Organizational Science. *Organizational Research Methods*, *21*(3), 525–547. https://doi.org/10.1177/1094428116677299

Vial, G. (2019). Reflections on quality requirements for digital trace data in IS research. *Decision Support Systems*, *126*(May), 113133. https://doi.org/10.1016/j.dss.2019.113133

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. *Communications of the Association for Information Systems*, *37*(AUGUST), 205–224. https://doi.org/10.17705/1CAIS.03709

Westen, D., & Rosenthal, R. (2003). Quantifying Construct Validity: Two Simple Measures. *Journal of Personality and Social Psychology*, *84*(3), 608–618. https://doi.org/10.1037/0022-3514.84.3.608

Xu, H., Zhang, N., & Zhou, L. (2020). Validity Concerns in Research Using Organic Data. *Journal of Management*, *46*(7), 1257–1274. https://doi.org/10.1177/0149206319862027

(*) Zhang, Z., Lee, H., Yoo, Y., & Choi, Y. T. (2022). Theorizing Routines with Computational Sequence Analysis: A Critical Realism Framework. *Journal of the Association for Information Systems, 22*(2), 589–630. https://doi.org/10.17705/1jais.00734