Association for Information Systems

AIS Electronic Library (AISeL)

Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023

General IS Topics

Dec 11th, 12:00 AM

A Framework for Developing Cross-Sectional Surveys

Maren Mehler
Technical University of Darmstadt, maren.mehler@tu-darmstadt.de

Sara Ellenrieder
Technical University of Darmstadt, ellenrieder@is.tu-darmstadt.de

Merve Turan Akdag
TU Darmstadt, merve.turan_akdag@tu-darmstadt.de

Amina Wagner
Technische Universität Darmstadt, wagner@is.tu-darmstadt.de

Izak Benbasat
University of British Columbia, izak.benbasat@sauder.ubc.ca

Follow this and additional works at: https://aisel.aisnet.org/icis2023

Recommended Citation

Mehler, Maren; Ellenrieder, Sara; Turan Akdag, Merve; Wagner, Amina; and Benbasat, Izak, "A Framework for Developing Cross-Sectional Surveys" (2023). *Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023*. 6.
https://aisel.aisnet.org/icis2023/generalis/generalis/6

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023 by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

A Framework for Developing Cross-Sectional Surveys

Completed Research Paper

Maren F. Mehler
Technical University of Darmstadt
Hochschulstraße 1, 64289 Darmstadt
mehler@is.tu-darmstadt.de

Sara Ellenrieder
Technical University of Darmstadt
Hochschulstraße 1, 64289 Darmstadt
ellenrieder@is.tu-darmstadt.de

Merve Turan Akdag
Technical University of Darmstadt
Hochschulstraße 1, 64289 Darmstadt
turan@is.tu-darmstadt.de

Amina Wagner
Technical University of Darmstadt
Hochschulstraße 1, 64289 Darmstadt
wagner@is.tu-darmstadt.de

Izak Benbasat
Sauder School of Business, University of British Columbia
2053 Main Mall, Vancouver, BC V6T 1Z2
izak.benbasat@sauder.ubc.ca

Abstract

Although the use of cross-sectional surveys is widespread in Information Systems (IS) research and related disciplines, few papers address the survey development process. In order to ensure a standardized approach, comparable and valid results, as well as to guide researchers in quantitative research methods, this paper presents a framework for the survey development process in IS. Based on a Design Science Research (DSR) methodology, the framework was derived from a structured literature review of leading IS journals and refined by three focus group discussions among IS experts. The framework includes several steps and considerations on the sample size, variable selection, their order in the survey, protection against bias, ensuring validity and reliability, and testing before administering the survey with a focus on documentation and reporting. Our framework supports quantitative research by providing a structured approach to create reliable and credible surveys.

Keywords: Cross-sectional surveys, survey development process, research methodology

Introduction

The Information Systems (IS) research domain highly values the accurate implementation and effective presentation of research methodologies. Ultimately, a sound research design and its execution are essential for any publication. Given the importance of methodology in IS research, a great body of literature exists that provides detailed guidance on how to design research projects and execute them properly. In this body of work, five overarching pillars of research methods and their respective guidance have emerged. Regarding literature reviews, Okoli and Schabram (2010), Schryen (2015), vom Brocke et al. (2009), and Webster and Watson (2002), for example, offer step-by-step instructions on how to perform literature reviews in a structured and reliable manner. For conducting qualitative research and its underlying grounded theory, Eisenhardt (1989), Myers and Newman (2007), or Urquhart et al. (2010) form the basis for almost every qualitative research project; independently of whether researchers are novices or experienced. Design science research (DSR) goes even further here, offering established papers that detail the steps a successful DSR project must go through to be relevant and rigor (e.g., Hevner et al., 2004; Kuechler & Vaishnavi, 2008; Peffers et al., 2007). Finally, Venkatesh et al. (2013), as an example, provide guidelines for mixed methods research in IS.

As a fifth research method, quantitative research, and especially cross-sectional survey studies, have been a key element of IS research for decades (Newsted et al., 1998). They allow to observe a phenomenon in a broad sample and to validate relationships statistically relatively inexpensive and easily if done right (Dinev et al., 2013; Melnyk et al., 2012; Recker, 2021). Surveys can be used for *exploration* – to become familiar with a topic; for *description* – to study behavior, opinions, processes or situations; or for *explanation* – to test theoretical and hypothetical relationships (Recker, 2021). To better define the scope of our framework, it is important to clarify our understanding of the term "survey" within the context of IS research. Compared to other quantitative methods, such as experiments, surveys are often administered to a broad sample, resulting in generalizable conclusions by using structured questionnaires (King, 2005; Mazaheri et al., 2020). In general, surveys can be categorized into cross-sectional and longitudinal surveys. The former is defined as a one-time data collection (Srivastava et al., 2015). Longitudinal surveys are characterized by surveying the same individuals at least twice, resulting in the measurement of differences over a longer period of time (Moorman, 2008). This paper refers primarily to cross-sectional surveys as they typically are part of almost every quantitative research and are regularly adopted by IS researchers. The data can originate from heterogeneous sources, nowadays especially online surveys (Che et al., 2022; Ciolkowski et al., 2003; Kranz, 2021; Mazaheri et al., 2020).

Despite its wide usage (Mazaheri et al., 2020), the literature on quantitative research (in IS) has primarily focused on two aspects: the development of constructs (e.g., Compeau et al., 2022; MacKenzie et al., 2011; Moore & Benbasat, 1991) and the statistical analysis of survey results, especially ensuring the quality of data through validity or reliability measures (e.g., Hair Jr. et al., 2017; King, 2005; Podsakoff et al., 2003; Schmitz et al., 2020), which is also the focus of the well-known website of Straub et al. (2022). However, the survey development process involves multiple steps, including constructing survey questions, designing the survey and administering the survey, before analyzing the data. To the best of our knowledge, no paper in IS has yet focused on "bridging the gap" between a researcher choosing to conduct a survey and the statistical analysis of survey data by providing detailed guidance on the steps involved in survey development, thus ensuring that the data collected is valid, reliable, and accurate. Although this process might be known by experienced researchers, providing an explicit framework can significantly improve data collection and therefore the quality of future research output. As such, the current state of the literature on quantitative research in IS remains insufficient in providing a comprehensive understanding of the research process, as the actual steps involved in the survey development process have received relatively little attention. In addition, despite surveys being used in various research fields, there also seems to be a dearth of research on the development of surveys. Since IS research is strongly methodology-driven, as many other fundamental methodology papers on, for example, structured literature review or DSR have been successfully published and IS is an interdisciplinary research area, we consider the development process of cross-sectional surveys in IS in this paper. Thus, the primary objective of this paper is to provide a framework consisting of a comprehensive set of guidelines to standardize the survey development process, thereby enhancing comparability and reproducibility across research. Especially for beginners in survey development, the framework offers a detailed and structured approach for conducting surveys, which ensures high-quality surveys that deliver accurate and reliable data. But although the concepts underlying the derived guidelines are probably known by advanced researchers, they may also benefit from this paper by using the proposed framework as a checklist that their surveys are developed to the highest standards. Finally, this paper aims to provide support to reviewers, offering a common framework for evaluating and providing better feedback to cross-sectional survey research papers. By approaching this broad range of researchers, we seek to support at all stages of the survey development process to enhance the rigor and quality of research.

We follow the DSR approach by Peffers et al. (2007) to derive a framework for the survey development process. To this end, we conduct a structured review of existing literature for the derivation of guidelines proposed in this paper to aggregate and synthesize the best practices currently available in IS research. The guidelines were first presented as part of a research project presentation to other IS researchers and then evaluated by three focus group discussions. They also expanded the framework with IS-specific tangible knowledge. Deriving the guidelines from the literature and then evaluating them through focus groups ensures that the framework is not only based on existing literature but is also practical and useful for researchers in the field of IS research. In addition, our guidelines are not intended to be prescriptive or restrictive. Instead, they provide a flexible framework that can be adapted to different research contexts and can be tailored to specific research questions that are suitable for quantitative research. For example, telephone- or face-to-face surveys can also benefit from these guidelines, but some of the guidelines may need to be adapted minimally for use in other types of survey approaches.

This paper is structured as follows: After the Introduction, which contains the problem statement and motivation of this paper as well as existing work, the Methodology is described, where the DSR approach, including the literature review and the focus groups, are described. This is followed by our framework in the results section. The paper concludes with a Discussion of contributions, limitations, and future research.

A Design Science Research Approach

To derive a structured framework for the standardized development of surveys that allows comparability and reproducibility, we chose a DSR approach similar to vom Brocke et al. (2009) or Nickerson et al. (2013). DSR allows us to iteratively derive guidelines first from literature and then improve them through a focus groups of IS researchers. Thus, we apply the DSR approach suggested by Peffers et al. (2007). First of all, following this well-established, concise approach ensures the rigorousness of the derived solution (Hevner et al., 2004). Second, while we aim to deepen our understanding of survey development in the IS research community, it is a key objective to ensure comprehensibility as well as broad practicability of the derived guidelines underlying our framework. As many surveys investigate real-world problems, the applied DSR methodology is suitable for developing an artifact in the form of a framework that allows for broad applicability (Baskerville et al., 2015; Venable & Baskerville, 2012). The derived framework defines clear guidelines for each step in the development process of a survey and thus provides a Level 2 contribution in terms of the DSR contribution types by Gregor and Hevner (2013). Example artifacts for Level 1 contributions are specific solution instantiations, while Level 2 contributions are, e.g., constructs, methods, models, design principles, and technological rules, and Level 3 contributions include the derivation of new design theories (Gregor & Hevner, 2013).

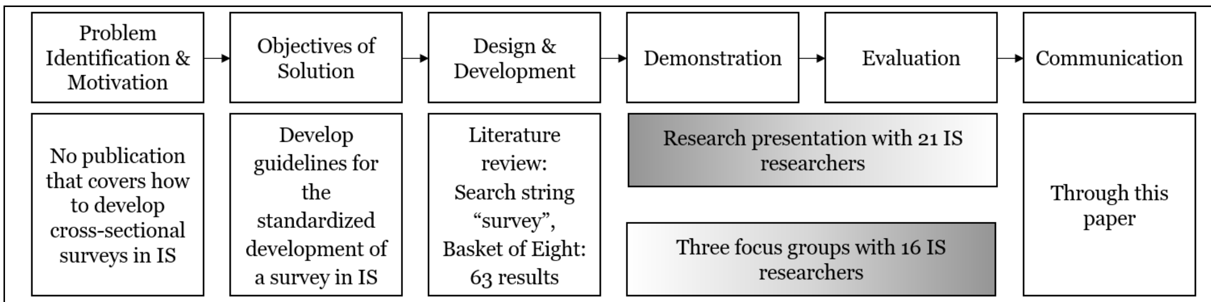

Figure 1. DSR Methodology for Our Study in Accordance with Peffers et al. (2007)

Figure 1 visualizes our approach, which consists of six distinct phases, namely 1) problem identification & motivation, 2) objectives of solution, 3) design & development, 4) demonstration, 5) evaluation, and 6) communication with possible starting points at 1) to 4) (Peffers et al., 2007). We began the design cycle in phase 3) by first reviewing existing literature on surveys in the Basket of Eight journals (*European Journal of Information Systems*, *Information Systems Journal*, *Information Systems Research*, *Journal of the Association for Information Systems*, *Journal of Information Technology*, *Journal of Management Information Systems*, *Journal of Strategic Information Systems*, and *MIS Quarterly*) to account for rigor in the design process (Hevner et al., 2004). A detailed description of the performed journal review is provided in the next section. Besides identifying the lack of clear guidance for developing surveys for phase 1, we were able to define the key objectives for the framework in process phase 2. Based on the surveys that were published in the Basket of Eight journals, guidelines for all process phases of the development phase of surveys were derived. Standardization and broad applicability of guidelines were the key objectives during this phase 3. Peffers et al. (2007) suggest demonstrating the artifact in phase 4 before evaluating it in phase 5. We demonstrate and subsequently evaluate the guidelines in different settings to ensure comprehensibility and effectiveness while also obtaining in-depth qualitative feedback from IS researchers. In a presentation of our research project with 21 IS researchers the identified research gap as well as derived guidelines were shown before an open discussion was conducted on the relevance, comprehensibility, and completeness of the framework. The main focus of the research presentation was to demonstrate the guidelines to a large group of researchers, who were then asked about previous their experiences with surveys and then were invited to participate in several focus groups according to their level of expertise. Focus groups allowed us to evaluate how selected groups of researchers with different prior experiences evaluate the derived guidelines (O'Nyumba et al., 2018). The DSR approach by Peffers et al. (2007) requires communication (phase 6) of the results obtained to appropriate audiences. As we primarily focus on surveys conducted in IS research, this paper allows us to present the conducted design cycle and final framework to IS researchers who can build their future research on the blueprint provided in this paper.

Literature Review

To derive the initial set of guidelines, we conducted a literature review following the five steps according to the approach suggested by vom Brocke et al. (2009). First, we **defined our review scope** using Cooper's (1988) Taxonomy which consists of six characteristics guiding researchers to specify their scope. Our *focus* is on research methods, as it is the case with quantitative surveys, and our *goal* is to integrate and combine the literature into a framework. We *organize* our review conceptually and take a neutral *perspective*. Our *audience* are both specialized and general scholars in IS and we provide representative *coverage* by focusing on articles in the Basket of Eight journals that ensures high quality as they have already passed an extensive review process. Second, we **conceptualized the topic**. We agreed on the most general search string "survey" to remain as broad as possible. Since we focus on cross-sectional surveys, we chose the following inclusion criteria: 1) At least one cross-sectional survey was conducted in the paper, and 2) the paper provides sufficient insights into the survey itself or its development process. We did not consider papers that focus on research methodologies for the literature review but read them to substantiate our framework. Third, in our **literature search**, we used the search string "survey" in WebOfScience which allowed us to specifically selecting only articles from the Basket of Eight journals as they are extensively reviewed and accepted as high-quality journals (Polites et al., 2012). This resulted in 741 articles. We read and applied the inclusion and exclusion criteria on the papers according to the publication year by first screening the titles and abstracts before reading them (Schryen, 2015). To reach saturation we **analyzed and synthesized** papers according to their publication year adding more papers until no paper provide any new information on the survey development process. Thus, we started with the 23 papers from 2022 by scanning the titles and abstracts. We excluded three papers that were not conducting a survey. The remaining 20 papers were then read carefully. Based on them, we developed a list of coding categories, such as used constructs, included control variables, measurements against biases, and types of testing the survey, which we then applied to the selected papers. Afterward, we used the categories to derive a first draft of guidelines. We already noticed saturation in the guidelines after analyzing the papers from 2022. We then extended our guidelines with articles from 2021, applying the same procedure as before, again excluding seven papers as they did not conduct a survey, resulting in 27 articles. As hardly any new aspects for the guidelines emerged, we chose as representatives two random articles from each of the last 10 years to finalize the guidelines. Finally, we used a backward search to include relevant methodology papers – which are not counted here as they are only used in the guidelines and not to derive them – and we did not perform a forward search as we only needed a representative set of survey papers. Overall, we included 65 (20 from 2022, 27 from 2021, 2 each from 2012-2020) papers in the literature review. A **research agenda** as the last step will be included in the discussion as an outlook on future research ideas.

Focus groups

To evaluate the framework, ensure the relevance and correctness of the guidelines, and for additional input, we conducted three focus groups. A qualitative methodology is beneficial, as the knowledge collected is not shared in public in the form of papers but is anchored in the interviewees based on their experience. The main advantage of focus groups compared to interviews is that a discussion takes place between the participants and different opinions influence the result (Gibson & Arnott, 2007). In addition, Gibson and Arnott (2007) recommend focus groups for evaluating the utility, quality, and efficacy in DSR. The three focus groups took place in March 2023 and consisted of five to six IS researchers who knew each other at least to some extent and with different research interest such as privacy, AI or healthcare. Six experts (one postdoc and five researchers at the end of their Ph.D. studies) with at least four (mean 10) already conducted surveys were invited to the first focus group, since we expected the highest amount of input here. The participants were asked to speak openly about their experiences and advice with surveys along the survey development process (see Figure 2) to improve and expand our guidelines. Following this, we asked five proficient researchers (in the middle stages of their Ph.D. studies) who had conducted at least one, but maximum two surveys to give input to each guideline by its title, hoping to gather in particular their difficulties and learnings to gain feedback on the usefulness, completeness and clarity of our guidelines. Finally, we evaluated the guidelines for ease of use, understandability and practicality with five inexperienced (no surveys conducted) researchers at the beginning of their Ph.D. studies by summarizing the guidelines. Between each focus group, the guidelines were adapted to the new insights. Following Tremblay et al. (2010), we conducted an exploratory focus group with the six experts ("E") and a confirmatory focus group with the five inexperienced participants ("I"). The focus group in between them consisted of five proficient participants ("P") was both exploratory and confirmatory. In order to maximize the feedback from the participants and to minimize the length of the interviews to avoid fatigue, we chose relatively small groups compared to Tremblay et al. (2010)'s suggestion of 6-12 participants. The same author moderated the focus groups and afterwards transcribed them for analysis. The discussions were, on average, 59 minutes. Finally, the focus groups were coded during content analysis using MAXQDA software (VERBI Software, 2021) by the authors to minimize the influence of individual researchers and to ensure objectivity. For this purpose, we applied the guidelines derived from the literature to the transcripts (deductive) and simultaneously identified new, context-specific aspects for the framework through an inductive approach (Mayring, 2000). From the focus groups we were able to refine the guidelines regarding tacit knowledge, for example how to order the questions in a survey or that it is important and normal to iterate back anytime during the development of a survey.

A Framework for Cross-Sectional Survey Development

In the following, the guidelines of the survey development process derived from the literature and enriched and validated by the focus groups are presented. In order to clearly distinguish the results of the literature search and the focus groups, the sources are always provided. As our focus is on the methodological part of a paper, the guidelines start after the research question has been defined and a theory – used synonymous with model or framework and describes the presumed relationships between the constructs – has been selected. The theory can be self-developed, e.g., derived from interviews, or it can be an established theory that has been extended by self-selected variables. Thus, the finding of a theory and the development of hypotheses are not included in this paper. However, especially guidelines 1, 2a, 3 and 4 (in Figure 2) are dependent on the chosen theory. In addition, as many papers exist on how to analyze data and this is depending on the specific research questions, our guidelines end before the analysis.

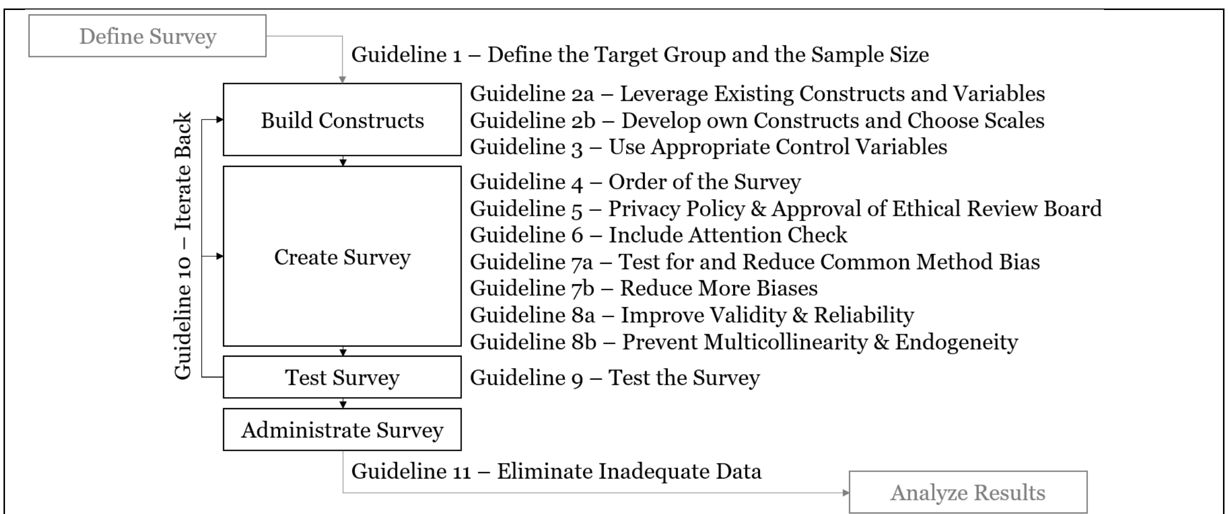

Figure 2. Framework for Cross-Sectional Survey Development

Figure 2 gives an overview of the framework, including the guidelines and their order for the development of a survey.

Guideline 1 – Define the Target Group and the Sample Size

First, the target group must be determined, because elements of other guidelines such as the type of attention check (see Guideline 6) or the phrasing of the items (see Guideline 2) are based on it (Expert 5 – E5). The target group depends on the research question and chosen theory and, in particular, three alternatives are used: A representative group (e.g., according to age and gender in a country (J. Wang et al., 2016)), a specific group (such as employees of companies (Kranz, 2021), managers of different companies (Liang et al., 2022), or members of a platform (Abhari et al., 2022)) or a student sample as a convenient sample (Serenko & Turel, 2021). In the IS literature, all three variants are used, but mainly specific groups.

The sample size ranges from 100 (Zhou et al., 2022) to 1281 (Rai et al., 2022) across the papers. Rules of thumb and formulas exist to calculate the minimum numbers; however, no paper includes them or any rationale on the number of participants. As the sample size is also depending on the requirements during the analysis, we will only give the suggestion to follow Maier et al.'s (2023) recommendation of the website from Wang and Ji (2020, http://riskcalc.org:3838/samplesize/) to calculate the sample size for cross-sectional surveys.

Furthermore, the researchers must consider the form of recruitment. Market research institutes are frequently approached for this purpose, as they offer representative samples and Steelman and Hammer (2014) did not find a statistical disadvantage in acquiring participants through platforms such as online crowdsourcing markets. However, these can quickly become very expensive (Steelman & Hammer, 2014). If a specific target group is being surveyed, researchers often have to recruit participants via contacts, e-mails, LinkedIn, etc. A student sample, on the other hand, is easy to recruit, as students can be asked to participate in courses, but has the most issues including low heterogeneity thus low representativeness (E1). In this context, also the question of compensation arises. At a market research institute, participants are often paid a fixed amount depending on the duration of the survey, a few dollars/euros. Students or members of platforms often receive an incentive in the form of a fixed amount or the prospect of a possible prize in gift cards or even products such as iPads to increase response rate (Feng et al., 2022; Melnyk et al., 2012). Employees or managers of companies often do not receive an incentive.

Therefore, we propose: Document **(Guideline 1.1 – G1.1)** the target group and why it was chosen. State **(G1.2)** how many potential participants exist, the sample size calculated and according to which formula or rule of thumb, how many participants have been contacted, and how many have participated in order to calculate the response rate. Specify **(G1.3)** which type of recruitment was chosen and **(G1.4)** whether incentives were offered.

Guideline 2a – Leverage Existing Constructs and Variables

From the chosen or developed theory measures for the survey must be derived. The existing, often abstract constructs from the theory are operationalized through specific variables (Lund Research Ltd, 2012). Overall two approaches exist: First, leveraging existing variables, including their items – variables are usually measured through multiple similar questions, which have the same scales, and second, developing new constructs and variables (see Guideline 2b). Relying on established constructs and variables has the advantage that their validity has already been proven and results can be compared across several papers (E1; E3; P1; P2; P3; P4; P5). If multiple variables exist for the selected construct, we recommend choosing the best known, most suitable and most recent one. Here, the relevance should be assessed (Compeau et al., 2022). Picking individual items from various existing variables is not recommended. Often, variables need to be adapted to the survey context. If only minor changes are needed, the items can be modified. For example, Cichy et al. (2021) chose the privacy concern collection measurement from Smith et al. (1996). Originally the item is "It usually bothers me when companies ask me for personal info." which is adapted to "It usually bothers me when the company asks me for personal driving data." The adaptation should not result in a completely new item (P4). To determine whether more substantial changes are required, Compeau et al. (2022) describe a four-step approach.

Existing (occasionally also self-created) items have to be translated sometimes. Based on Brislin (1970), the following procedure has been used by many researchers (e.g., Maier et al., 2021): The questions are translated into the target language by a bilingual individual and then translated back by a second bilingual individual. If both versions in the original language are now considered to be without differences in meaning, a pretest should be carried out in the target language. Afterward, the questions should be re-evaluated by bilingual persons, one group seeing only the original, one group seeing the translation, and one group seeing both. In case no differences in the results are detected, the translation can be adopted (Brislin, 1970). This process is often carried out in a condensed version, for example by relying on professional translators to translate the text into the target language and then re-translating the text back into the original language (e.g., Hovav et al., 2021; L. Wang et al., 2022; E1).

Therefore, we propose: Document **(G2a.1)** the source of constructs and variables, **(G2a.2)** their relevance, e.g., following Compeau et al. (2022), **(G2a.3)** the extent to which they have been adapted to the new context and **(G2a.4)** how there were translated. Best practice is to report all constructs, variables and their items in the paper.

Guideline 2b – Develop Own Constructs and Choose Scales

Adequate constructs and variables may not always exist in the literature. However, developing constructs should be the exception as it is a time-consuming process (E2; E3; E5). The development of new constructs usually follows MacKenzie et al.'s (2011) ten steps scale development process from the definition of the construct over the generation of items, the collection of data for testing, the elimination of items, and the validation of the development of norms (e.g., Hoehle & Venkatesh, 2015; Riemenschneider & Armstrong, 2021). They also elaborate on the difference between formative and reflective measurement models and thus provide a solid guideline for the development of constructs.

Variables can consist of multiple (preferred) or a single item, but can be also answered through open-ended, ranking or many more types of questions. Churchill (1979) recommended the use of multiple items because this results in higher reliability (see Guideline 8a) and is suitable for structural equation modeling. But it is also possible to use single items, often done by practitioners because of their simplicity. They are accepted in marketing research if they are doubly concrete constructs (Bergkvist & Rossiter, 2007). To measure items, in IS and related areas, mostly Likert scales have been adopted. Here, an item is given as a statement and the respondents have to give their attitude on a range from "strongly disagree" over "neutral" to "strongly agree" (Flamer, 1983; Likert, 1932). Commonly the range has seven steps (e.g., Cichy et al., 2021; Lin et al., 2021), but also five (e.g., Ke et al., 2021) and ten (e.g., Rai et al., 2022) are used. Regarding the choices of the best length and whether the scales should be even or odd, no clear direction is given (Leung, 2011). For the length, the literature argues in favor of long scales because of the larger variety of options for the participants so they are supposed to find their true value more likely (Joshi et al., 2015). On the other hand, an overly fine scale can be overwhelming (Matell & Jacoby, 1971). Sometimes it can be beneficial to change the wording and use semantic differential scales (Chin et al., 2008). This results in more variable and flexible scales and shorter items, e.g., instead of "Using the system enhances my effectiveness" on a scale from strongly disagree to strongly agree the item would be "This system is …" ranging from efficient to inefficient (Chin et al., 1996, p. 690). In addition, there are other scales that can be used, such as the Stapel scale or the Thurstone differential scale (Albaum, 1997). If an item comes with a scale, we recommend using it. Other types of questions are depending on the variable. Open-ended questions can, for example, be used to find reasons in participants behavior and analyzed qualitative or ranking, for example, in conjoint analyses.

Therefore, we propose: Document **(G2b.1)** which constructs are newly developed and **(G2b.2)** the process of their development. In addition, we recommend describing this process in a new paper or at least as detailed as possible including testing of the new constructs in the current paper. Additionally, **(G2b.3)** list which scale was used for all variables.

Guideline 3 – Use Appropriate Control Variables

Control variables or if they refer to a person demographic data are included to account for effects that may confound the relationship between independent and dependent variables and thus to rule out alternative explanations (Atinc et al., 2012). This also allows identifying potential group differences (Kung et al., 2015). Furthermore, demographic data can be used to determine whether certain quotas are already met, e.g., age distribution in the country under consideration (E1; P1).

If the survey focuses on individuals, in particular, age and gender are included in the survey, as well as income, employment status, work experience, and sometimes marital status and the number of children.

Depending on the topic, more specific variables can be utilized (e.g., Che et al. (2022) measured the experience in online shopping). Demographic data is often presented in the methodology section. When collecting demographic data, privacy must be ensured to avoid individuals being identifiable (see Guideline 5). For surveys focusing on companies, control variables such as size and industry are measured in particular. In addition, as such surveys require one person to respond on behalf of the company, demographic data for this person is often included as well *"because the demographic data are always collected"* (E3; P1; P3; P4). However, measuring every control variable that might have an influence is not the solution. Carlson and Wu (2012) even recommend omitting control variables if there is a lack of evidence to support their inclusion. Bernerth and Aguinis (2016) presents a decision tree: Only if the variable is likely to have an influence that has already been examined and confirmed by other researchers and offers an alternative explanation, the inclusion of the control variable is reasonable. However, the focus group participants stated that they prefer to measure more control variables than to realize later that they missed one (P1; P4).

Therefore, we propose: Document **(G3.1)** all control variables and demographic data that are collected during the survey. Include if possible **(G3.2)** a valid rationale why the control variable is measured and **(G3.3)** show an overview of the values and their influence on the model in the analysis.

Guideline 4 – Order of the Survey

After defining the measurements, they must be arranged within the questionnaire "*for the quality (...) of the survey*" (E6; Drury & Farhoomand, 1997). Regarding the order of the survey, no explicit statements can be found in the examined literature. A survey should start with a privacy policy focusing on confidentiality and anonymity and information about the survey (see Guideline 5). In addition, important definitions should be included at the beginning to achieve a common understanding by using not only plain text, but also images (Kranz, 2021; P4). The variables can be structured based on the relationship between the constructs from in the theory, placing mediators at the end (E6; P2; P5). However, sometimes it is useful to measure the dependent variable first to avoid influence by other questions (e.g., questions about actual behavior can be influenced if people are asked how they handle privacy) (E1; P3). Regarding control variables, they are often included as first (e.g., Wiener et al., 2021), which makes it possible to check directly if quotas (e.g., balanced sample size according to gender) have to be met (P3), or last part of the survey (e.g., Abhari et al., 2022), "*because it's something that tends to bore the participants*" (P3). Finally, it is useful to offer an open feedback box in the end (E2).

In online surveys, the number of questions on a page should be limited so that participants only have to scroll once (P1; P4). Each page should consist of approximately the same number of questions or take the same amount of time (P2; P3). In addition, a page should be thematically independent (E5; P1). A progress indicator helps the participants to maintain an overview and remain motivated (P3; P4). Furthermore, a back button should not be included if the participant should not change their original opinion (E5; P1; P2). Moreover, making questions mandatory has the advantage that users must not be eliminated at the end because of missing data (see Guideline 11), but should be avoided with private questions (P1; P2). Finally, it is useful to block questions of one variable with Likert scales in a matrix (P1), but reversed questions in a matrix should be placed carefully to avoid that participants overlooked them (E2; E4). Finally, the questions should also not be overly complex and long (P3).

At the variables' and items' levels, a further question arises whether they should be asked in the same order or randomly for each participant as it can influence the reliability and construct validity (see Guideline 8a). To reduce common method bias, a recommendation is to randomize the order of the survey questions (Podsakoff et al., 2003; Sharma et al., 2022; L. Wang et al., 2022). Wilson et al. (2021) recommend not to use intermixed approaches (mixing items from different variables), but to order the variable blocks randomly.

In this context, the length of the survey needs to be briefly addressed. Although no specific rules exist, the length of the survey has an impact on the response rate and can therefore threaten the validity of the survey (Chin et al., 2008; Melnyk et al., 2012). Additionally, the reliability of later questions may decline as participants become increasingly fatigued. The experts from our focus groups recommend a survey length of 10 to 20 minutes.

Therefore, we propose: Document **(G4.1)** in which order the survey components were arranged. If an online survey was conducted, **(G4.2)** describe the settings in the tool used to set up the survey. If the order of questions is random, **(G4.3)** describe how it was set for different participants, and indicate **(G4.4)** how long the survey takes on average.

Guideline 5 – Privacy Policy & Approval of Ethical Review Board

To ensure privacy, it is important to include information about the storage and use of the collected (demographic) data for the participants and inform them about their rights (e.g., GDPR and that the survey is voluntary so they can leave anytime) (P1; P2). Often, universities provide information on how to implement privacy policies on their website (P3). Especially in psychology, ethical principles are ensured by the approval of an ethical review board. This should be considered depending on the topic and questions, e.g., when collecting sensitive such as medical data from the participants.

Therefore, we propose: Document **(G5.1)** if a privacy policy is provided and **(G5.2)** if an ethical board approves the survey.

Guideline 6 – Include Attention Checks

Since participants may become inattentive during surveys, we recommend including at least one attention check to improve data quality. A specification on which attention check was included was rarely provided in the literature. Only the following example was found: Participants are asked to “please select strongly disagree” (Wiener et al., 2021).

Abbey and Meloy (2017) provide an overview of different types of attention checks along with the simplicity of implementation and objectivity of these checks: Logical statements, directed queries (including the example above), manipulation checks, open-ended queries, infrequency, response pattern/ time, honesty check, reverse scaling, memory recalls, and outlier detection. Except for response pattern/ time and outlier detection, attention checks must already be planned during survey creation. An attention check should be placed in the middle of the survey (P4). It should be noted that if you ask a “Please check strongly disagree” question, it is possible that the questions surrounding it will be influenced by the strong statement (E3). The advantage of attention checks is that they can be used to easily eliminate data sets where participants have made little effort (see Guideline 11), thus ensuring high data quality and that the survey reflects the true behaviors, attitudes, and beliefs of the participants (Abbey & Meloy, 2017). Especially in the context of online surveys using tools like Amazon Mechanical Turk, the inclusion of attention checks is recommended (Paas & Morren, 2018). However, Hauser and Schwarz (2016) discovered that Amazon Mechanical Turk respondents learned how to answer these attention checks.

Therefore, we propose: Document **(G6.1)** which type of attention check was used and **(G6.2)** how many participants failed it and were thus eliminated.

Guideline 7a – Test for and Reduce Common Method Bias

Often, surveys are checked for the common method bias (CMB or common method variance) (Fuller et al., 2016; Podsakoff et al., 2003). CMB can occur when data are collected from the same source. A faulty measurement caused by a bias can lead to incorrect results and thus flawed conclusions. Different causes for the CMB exist: “a common rater, a common measurement context, a common item context, or from the characteristics of the items themselves” (Podsakoff et al., 2003, p. 885). An overview of options for control in design or in statistics, i.e., after the survey has been conducted is provided by Podsakoff et al. (2003). The focus here will be on the options in the design: Whenever possible, the independent and dependent variables should be collected from different sources in order to minimize the common influence on them, but collecting data from the same respondents does not automatically result in CMB. When conducting longitudinal surveys, it is possible to query the two types of variables in different surveys (Gong et al., 2021; L. Wang et al., 2022). In case it is not possible to collect the data from different sources, a separation – temporal by ordering the survey questions (see Guideline 4), psychological by cover story, methodological or proximal by different response formats – can be arranged. Another possibility to reduce CMB is to guarantee anonymity to the participants (see Guideline 5) and to assure that there are no right or wrong answers. Especially to reduce priming effects it is recommended to vary the order of questions. Using a marker variable (e.g., “Coffee is important in my life”) in the research model could also support determining CMB. If no high correlation with other variables can be found, the possibility for CMB is low (Maier et al., 2021; E2; E3; P3). Finally, by ensuring high quality constructs and testing them, CMB can be reduced (P3) (see Guideline 2). Cram et al. (2022, p. 440f.) summarize “keeping questions simple, focused, and concise; avoiding double-barreled questions and conceptual dependence between dependent and independent variables; using randomized items and reiterating respondent anonymity along with the exclusive research purpose of our study". This is extended by Jordan and Troth (2020) with higher item clarity and including reverse items or balancing positive and negative items.

After the data collection, the data can be additionally tested for CMB. As these tests are usually part of the analysis rather than the methodology, only two will be mentioned and briefly described here (Maier et al., 2021): First, Harman's single factor test, which indicates whether the majority of the variance can be explained by one single factor. Second, extremely high correlations ($r > 0.90$) are an indicator of CMB.

Therefore, we propose: Document **(G7a)** how the survey was protected against common method bias.

Guideline 7b – Reduce Additional Biases

Surveys are vulnerable to other types of biases as well. Therefore, again, the survey should be protected against certain biases in the survey design:

Social desirability bias: When the survey is distributed, social desirability bias can be reduced by assuring anonymity (see Guideline 6) or by sharing the survey on a survey website (Gong et al., 2021).

Non-response bias (or response bias): When conducting surveys, usually not all of the requested participants respond, which results in a subsample of the total sample. Those who do not participate may differ from the responding participants in important aspects. Especially the attractiveness of the study can increase the number of respondents (E4). Also, Armstrong and Overton (1977, p. 396f.) suggest the following options: Comparison with known values for the population, subjective estimates, or extrapolation methods. However, the literature shows mostly a-posteriori methods as test for significant differences using t-tests between early respondents and late respondents or ANOVA on the measurements and control variables (e.g., Kranz, 2021). Furthermore, sometimes it is the intention that only certain groups participate: *"It's completely reasonable that some groups won't be included. If you're measuring AI acceptance, it doesn't make much sense to ask people who don't use the Internet, because they'll never be in touch with AI"* (E2).

Other biases, such as self-selection bias, exist, however, we could not find protections against them.

Therefore, we propose: Document **(G7b)** how the survey was protected against the respective biases.

Guideline 8a – Improve Validity and Reliability

Similar to the common method bias, validity and reliability are often checked after the data has been collected, but some approaches exist to ensure them while implementing the survey. For more information on validity and reliability, we recommend Straub et al. (2022).

Validity is achieved if a measure is equal to the true value (Churchill, 1979). Different types of validity can be found in the literature: Conclusion validity indicates the extent to which a relationship between two constructs is random (Schmitz et al., 2020). Construct validity indicates to what extent the construct measures what it is supposed to measure (Peter, 1981; Voorhees et al., 2016). Content validity indicates the extent to which the variable captures what it is supposed to measure (Schmitz et al., 2020). Convergent validity indicates the extent to which the items of a variable positively correlate and discriminant validity if a variable is empirically distinct from different variable (Hair Jr. et al., 2017). External validity indicates to what extent the results are generalizable "across different measures, persons, settings, or times" (King, 2005, p. 882). In general, surveys provide high external validity (Ciolkowski et al., 2003) and are therefore easily generalizable. Internal validity indicates to what extent causality is present (Schmitz et al., 2020).

When creating the survey, the following is important to improve validity: Internal validity can be ensured by using well-established measurements and collecting data carefully (Maier et al., 2021; P2; P3). Content validity requires careful development of the constructs, especially by using card or Q sort during the scale development (Moore & Benbasat, 1991). In addition, it can be enhanced by pre-testing the instruments and receiving professional advice from experts (Hovav et al., 2021; Schmitz et al., 2020). Construct validity can also be increased by testing the survey (see Guideline 9). Many tests and procedures exist that can be performed a posteriori to check if the measures are valid. Such as the average variance extracted (AVE) for convergent validity of reflective constructs with a threshold of 0.50 (see Fornell & Larcker, 1981; Hair Jr. et al., 2006, 2017; and e.g., Srivastava et al., 2015; Zhou et al., 2022). For discriminant validity first the cross-loadings can be compared, where higher loadings for another variable would violate validity (e.g., Abhari et al., 2022); second the Fornell-Larcker criterion where the square root of AVE should be higher than the variables' correlations (e.g., Srivastava et al., 2015) or third – as both measures have weaknesses – the heterotrait-monotrait ratio (HTMT) where the value should be below the threshold of 0.90 to suggest discriminant validity (Hair Jr. et al., 2017; Henseler et al., 2015; e.g., Maier et al., 2021).

Reliability "describes the extent to which a measurement variable or set of variables is consistent in what it is intended to measure across multiple applications of measurements" (Straub et al., 2022). To ensure reliability prior to the data collection, only one action could be identified: Use multi item variables (Bergkvist & Rossiter, 2007; Churchill, 1979). To test reliability of the variables after data collection in general two measures are used: Cronbach's alpha with a threshold of 0.70 (see (Hair Jr. et al., 2006) – (e.g., Cichy et al., 2021; Dinev et al., 2013; Hoehle & Venkatesh, 2015) or composite reliability (CR) also with a threshold of also 0.70 (see (Hair Jr. et al., 2006) and (e.g., Abhari et al., 2022; L. Wang et al., 2022)). Both indicate internal reliability if the value exceeds the threshold, but Cronbach's alpha is more likely to underestimate and CR to overestimate the internal reliability (Hair Jr. et al., 2017). Therefore, it is best to calculate and report both.

Therefore, we propose: Document **(G8a.1)** how validity and **(G8a.2)** reliability were ensured, improved and tested. Be careful to differentiate between reflective and formative constructs.

Guideline 8b – Prevent Multicollinearity and Endogeneity

Multicollinearity indicates a correlation between several variables. Multicollinearity is problematic for formative constructs, but desired for reflective constructs (Petter et al., 2007). When selecting variables (see Guideline 2), it is important to consider whether there is conceptual redundancy. Multicollinearity can be determined afterwards, for example, by calculating the variance inflation factor (Lin et al., 2021; Maier et al., 2021). Here the rule of thumb is that if the factor is > 5 moderate or > 10 high multicollinearity is present (see Craney & Surles (2002)).

Endogeneity results if an independent variable and the regression's disturbance term are correlated which can bias the collected data and reduce validity (Sande & Ghosh, 2018). Endogeneity arises from measurement errors, simultaneity, or omitted variables (Sande & Ghosh, 2018). The following four options are recommended to address endogeneity: First, control variables (see Guideline 3) can reduce bias due to omitted variables. Second, when possible, binary variables can also be adopted as independent variables instead of Likert scales. Third, collect instrumental variables from other data sources if common method bias is the cause of endogeneity (see Guideline 7a). Fourth, measure the instrumental variables at an earlier time point than the endogenous independent variable if simultaneity is the issue (see Guideline 4) (Sande & Ghosh, 2018). From the literature review a Durbin-Watson test (e.g., Rai et al., 2022) is one approach to determine endogeneity after data collection. Others can be found in Sande and Ghosh (2018).

Therefore, we propose: Document **(G8b.1)** how multicollinearity and **(G8b.2)** endogeneity were prevented.

Guideline 9 – Test the Survey

Every survey must be tested to identify flaws before administering, as small errors can quickly occur, such as a question being phrased unclear in meaning, because it was translated incorrectly. In addition, the flow, order of the questions (see Guideline 4), and timing are also factors that can be improved with testing a survey (Bolton, 1993). When preparing a pretest, five steps are stated by Hunt et al. (1982): First, determine which parts of the survey should be tested. We recommend testing the whole survey. Second, the method for the pretest must be determined. Depending on the type of pretest, it can be performed qualitatively or quantitatively (Kranz, 2021). Third, it must be specified who will conduct the pretest. Fourth, decide who will participate in the pretest. The following procedures were found in the reviewed literature: (1) Pretest with colleagues, who are not part of the research team, but experts in the research area (e.g., Cram et al., 2022; Wiener et al., 2021). Here, especially the wording and survey flow (Wiener et al., 2021), readability, understandability, and realism of the instrument (Trinkle et al., 2021) can be adjusted. (2) Pretest with students as a convenience sample to check the validity and reliability of the scales (e.g., Serenko & Turel, 2021; Sharma et al., 2022; Trinkle et al., 2021; L. Wang et al., 2022). (3) Pretest with potential participants (e.g., Hovav et al., 2021), for example, with a purchased sample (Seymour et al., 2021). The focus group participants preferred a combination of academic experts followed by potential participants (E1; P1; P2; P4). And fifth, how large the sample should be. Depending on the previous four steps and the sample size, a pretest sample between 5 – for feedback from researchers (e.g., Cram et al., 2022) or qualitative feedback (e.g., Kranz, 2021) – and 50 participants for a more convenient sample (e.g., L. Wang et al., 2022) or for the target group of the main survey (e.g., Abhari et al., 2022), should be sufficient (Hunt et al., 1982).

Quickly forgotten during testing is to ensure that basic aspects such as all branches are functioning (P1), the survey works on every common device, multiple languages are equivalent (P2), but also that the variables are properly implemented (E1; E2; E6). This can be achieved for example via a comment function of the tool (E1; P2) or open feedback in the survey (E6). Furthermore, self-developed constructs should always be tested extensively (see Guideline 2b).

Therefore, we propose: Document **(G9.1)** which part of the survey was tested and **(G9.2)** how. Specifically address **(G9.3)** which test group was contacted, **(G9.4)** how many participants tested the survey, **(G9.5)** in which order and **(G9.6)** with which focus. In addition, **(G9.7)** provide information on the changes after each test.

Guideline 10 – Iterate Back

In particular after testing the survey, but also any time before administering the survey, it is possible to iterate back and adjust previous decisions, e.g., to adapt items if they are difficult to understand (P4). Iterating back into the survey development is not a one-time step, but should be repeated several times. P2 believes this guideline is the most important: "*You should take as much time as possible to optimize the survey before you start to collect the data, because it hurts less than if you haven't done it and then collected data that you can't use because you have to throw everything away and then start all over again.*" (P2).

Therefore, we propose: Document **(G10)** what adjustments were made when iterating back in the survey development process.

Guideline 11 – Eliminate Inadequate Data

Some data sets need to be removed to avoid altering the results of the survey, since surveys are prone to human mistakes. The following data sets can be removed or are removed automatically by survey tools (e.g., Lin et al., 2021): Data sets that are incomplete (e.g., if single items have not been answered (P2)), where the attention check was failed (see Guideline 6) (P4), that do not meet a requirement (e.g., Riemenschneider and Armstrong (2021) eliminated 29 data sets of individuals who are not in the IS field, as this is the target participant group; however, it is recommended to sort these persons out with a question at the beginning), containing inconsistent information (e.g., years in IT > age (Riemenschneider & Armstrong, 2021)), where the response time is too short or too long (e.g. Wiener et al. (2021) sorted out if the minimum time (based on preliminary testing) is not met (L. Wang et al., 2022; E6)), or that have been identified as so-called straight-liners (P3). This describes data sets where the participant almost always gives the same answer on the same response scale. All previous types of outliers can be summarized under "error outliers" which can be removed or corrected. However, some outliers might also be authentic answers which provide important or unexpected insights which should not be removed but studied (Aguinis et al., 2013; P4). For an overview of approaches for identification and handling, see Aguinis et al. (2013) and Kim et al. (2019).

Therefore, we propose: Document **(G11)** which data sets are eliminated and why by defining exclusion criteria (P2) and handling strategies to ensure transparency.

Contributions, Limitations and Future Research

This paper presents a framework consisting of 11 guidelines which describe the survey development process. The guidelines begin after deciding on a theory to be evaluated and include choosing variables, arranging them, ensuring quality, testing the survey, and end before the data is analyzed. A focus is on the information that should be provided when writing the paper. This framework was derived from a structured literature search among the papers of the last ten years in the Basket of Eight journals to ensure high quality in the IS field and validated by three focus groups. Thus, we combine existing knowledge to create a standardized process for survey development. The framework is addressed to beginners who need support in developing a survey, but also more experienced researchers can benefit from a standardized approach, as well as to experts, as open issues regarding the survey development process are identified, and to reviewers in order to give them an overview of the most important aspects of a survey for their feedback. To ensure this, the three focus groups were conducted among experts, proficient, and inexperienced researchers, who all agree that this framework will be helpful in the development process to have an "overview" (E6) to "go through the process from front to back" (I4) and "having the requirements for a survey all at a glance, so that you don't have to gather them from different sources" (P1).

We provide several **contributions for IS research**, especially with regard to quantitative research methodology. First, with our framework, we present a development process that can be applied to the creation of surveys and thus lead to high quality surveys by standardizing them in order to achieve comparability and reproducibility. This is increased at the same time through the focus on the documentation. Thus, we especially enable novice researchers to face the challenge of quantitative data collection and to deliver rigor results. Second, our framework provides a structure that can be used to standardize other methodologies such as experiments or longitudinal surveys. Although DSR is most commonly used to derive design for (IT) artifacts, we adopt it to derive our framework. This is common in research on methodologies (e.g., Nickerson et al., 2013; vom Brocke et al., 2009). Thus, we would like to emphasize with our paper that the DSR approach is very versatile. Third, while our framework focuses primarily on cross-sectional survey development, it can be applied to longitudinal surveys and to some extent to experiments. However, they may not be suited for published statistics because we prioritize the design and data collection phase of survey development and neither cover the process of analyzing the collected data set nor reporting of the results. In addition, we assume that the framework is to some extent transferable to related research areas. The framework for the survey development process can also be utilized in **practice**, as surveys are regularly conducted here as well, which will be of higher quality after applying the guidelines. This could also lead to surveys created in practice being interesting and relevant for research.

Our paper is subject to some **limitations**. As we focus our literature search on high quality papers from the Basket of Eight journals, we exclude a large number of papers that are published at conferences. However, since the aim is to provide the most comprehensive framework possible, we assume that more of the survey development process is documented and reported in the considered papers. In addition, the number of participants in the focus groups is limited, but we tried to include different levels by dividing them according to their knowledge. Finally, our framework cannot guarantee internal validity, even if all guidelines are followed as this also depends on the selection of the theory and the analysis and interpretation of the results. But our framework is not intended to be prescriptive or limiting. Since IS is a very interdisciplinary field, we cannot present all possible issues. Moreover, to some aspects such as the length of the survey, there is no consensus neither in the literature nor among the participants of the focus group. This leads to **future research**. For example, open topics like how to decide for a theory, how to derive hypotheses, the number of participants depending on the sample and the outlet, especially also for pretesting, the optimal length of a survey, or additional methods to prevent biases or ensure validity and reliability during the survey development process should be explored. Also, a comparison between the requirements of a survey and the final implementation and reporting is interesting. We will follow up the paper with a validation of the framework by conducting a sample survey where all guidelines are applied and described in detail.

Acknowledgements

This research and development project is/was funded by the German Federal Ministry of Education and Research (BMBF) within the "Innovations for Tomorrow's Production, Services, and Work" Program (funding number 02L19C150) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by Design: Using Attention Checks to Detect Inattentive Respondents and Improve Data Quality. *Journal of Operations Management*, *53–56*, 63–70.

Abhari, K., Davidson, E., & Xiao, B. (2022). Inventing Together: The Role of Actor Goals and Platform Affordances in Open Innovation. *Journal of the Association for Information Systems*, *23*(1), 264–302.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). *Best-Practice Recommendations for Defining, Identifying, and Handling Outliers*. *16*(2), 270–301.

Albaum, G. (1997). The Likert Scale Revisited: An Alternate Version. *Journal of the Market Research Society*, *39*(2), 331–348.

Armstrong, J. S., & Overton, T. S. (1977). Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research*, *14*(3), 396.

Atinc, G., Simmering, M. J., & Kroll, M. J. (2012). Control Variable Use and Reporting in Macro and Micro Management Research. *Organizational Research Methods*, *15*(1), 57–74.

Baskerville, R. L., Kaul, M., & Storey, V. C. (2015). Genres of Inquiry in Design-Science Research. *MIS Quarterly*, *39*(3), 541–564.

Bergkvist, L., & Rossiter, J. R. (2007). The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs. *Journal of Marketing Research*, *44*(2), 175–184.

Bernerth, J. B., & Aguinis, H. (2016). A Critical Review and Best-Practice Recommendations for Control Variable Usage. *Personel Psychology*, *69*, 229–283.

Bolton, R. N. (1993). Pretesting Questionaires: Content Analyses of Respondent's Concurrent Verbal Protocols. *Marketing Science*, *12*(3), 280–303.

Brislin, R. W. (1970). Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216.

Carlson, K. D., & Wu, J. (2012). The Illusion of Statistical Control: Control Variable Practice in Management. *Organizational Research Methods*, *15*(3), 413–435.

Che, T., Peng, Z., Lai, F., & Luo, X. (2022). Online Prejudice and Barriers to Digital Innovation: Empirical Investigations of Chinese Consumers. *Information Systems Journal*, *32*(3), 630–652.

Chin, W. W., Johnson, N., & Schwarz, A. (2008). A Fast Form Approach to Measuring Technology Acceptance and Other Constructs. *MIS Quarterly*, *32*(4), 687–703.

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (1996). A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results From a Monte Carlo Simulation Study and Voice Mail Emotion/Adoption Study. *ICIS 1996 Proceedings*, *14*(2), 21–41.

Churchill, G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, *16*(1), 64.

Cichy, P., Salge, T. O., & Kohli, R. (2021). Privacy Concerns and Data Sharing in the Internet of Things: Mixed Methods Evidence from Connected Cars. *MIS Quarterly*, *45*(4), 1863–1892.

Ciolkowski, M., Laitenberger, O., Vegas, S., & Biffl, S. (2003). Practical Experiences in the Design and Conduct of Surveys in Empirical Software Engineering. In *Empirical Methods and Studies in Software Engineering* (pp. 104–128).

Compeau, D., Correia, J., & Thatcher, J. B. (2022). When Constructs Become Obsolete: A Systematic Approach to Evaluating and Updating Constructs for Information Systems Research. *MIS Quarterly*, *46*(2), 679–711.

Cooper, H. M. (1988). Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews. *Knowledge in Society*, *1*(1), 104–126.

Cram, W. A., Wiener, M., Tarafdar, M., & Benlian, A. (2022). Examining the Impact of Algorithmic Control on Uber Drivers' Technostress. *Journal of Management Information Systems*, *39*(2), 426–453.

Craney, T. A., & Surles, J. G. (2002). Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering*, *14*(3), 391–403.

Dinev, T., Xu, H., Smith, J. H., & Hart, P. (2013). Information Privacy and Correlates: An Empirical Attempt to Bridge and Distinguish Privacy-related Concepts. *European Journal of Information Systems*, *22*(3), 295–316.

Drury, D. H., & Farhoomand, A. (1997). Improving Management Information Systems Research: Question Order Effects in Surveys. *Information Systems Journal*, *7*, 241–251.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *Academy of Management Review*, *14*(4), 532–550.

Feng, Y. (Katherine), Claggett, J. L., Karahanna, E., & Tam, K. Y. (2022). A Randomized Field Experiment to Explore the Impact of Herding Cues as Catalysts for Adoption. *MIS Quarterly*, *46*(2), 1135–1163.

Flamer, S. (1983). Assessment of the Multitrait-Multimethod Matrix Validity of Likert Scales via Confirmatory Factor Analysis. *Multivariate Behavioral Research*, *18*(3), 275–306.

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models With Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50.

Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common Methods Variance Detection in Business Research. *Journal of Business Research*, *69*(8), 3192–3198.

Gibson, M., & Arnott, D. (2007). The Use of Focus Groups in Design Science Research. *ACIS 2007*

Proceedings, 327–337.
Gong, X., Cheung, C. M. K., Zhang, K. Z. K., Chen, C., & Lee, M. K. O. (2021). A Dual-Identity Perspective of Obsessive Online Social Gaming. *Journal of the Association for Information Systems*, *22*(5), 1245–1284.
Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337–355.
Hair Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th ed.). Pearson Prentice Hall.
Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd ed.). SAGE Publications Inc.
Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: Mturk Participants Perform Better on Online Attention Checks Than Do Subject Pool Participants. *Behavior Research Methods*, *48*(1), 400–407.
Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A New Criterion for Assessing Discriminant Validityin Variance-Based Structural Equation Modeling. *Journal of the Academy of Marketing Science*, *43*, 115–135.
Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Research. *MIS Quarterly*, *28*(1), 75–105.
Hoehle, H., & Venkatesh, V. (2015). Mobile Application Usability: Conceptualization and Instrument Development. *MIS Quarterly*, *39*(2), 435–472.
Hovav, A., Gnizy, I., & Han, J. (2021). The Effects of Cyber Regulations and Security Policies on Organizational Outcomes: A Knowledge Management Perspective. *European Journal of Information Systems*, 1–19.
Hunt, S. D., Sparkman, R. D., & Wilcox, J. B. (1982). The Pretest in Survey Research: Issues and Preliminary Findings. *Journal of Marketing Research*, *19*(2), 269.
Jordan, P. J., & Troth, A. C. (2020). Common Method Bias in Applied Settings: The Dilemma of Researching in Organizations. *Australian Journal of Management*, *45*(1), 3–14.
Joshi, A., Kale, S., Chandel, S., & Pal, D. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology*, *7*(4), 396–403.
Ke, W., Kang, L., Tan, C. H., & Peng, C. H. (2021). User Competence With Enterprise Systems: The Effects of Work Environment Factors. *Information Systems Research*, *32*(3), 860–875.
Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2019). Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail – Web Mixed-Mode Surveys. *Social Science Computer Review*, *37*(2), 214–233.
King, W. R. (2005). External Validity in IS Survey Research. *Communications of the Association for Information Systems*, *16*(1), 880–894.
Kranz, J. (2021). Strategic Innovation in IT Outsourcing: Exploring the Differential and Interaction Effects of Contractual and Relational Governance Mechanisms. *Journal of Strategic Information Systems*, *30*(1), 101656.
Kuechler, W., & Vaishnavi, V. (2008). On Theory Development in Design Science Research: Anatomy of a Research Project. *European Journal of Information Systems*, *17*(5), 489–504.
Kung, L., Cegielski, C. G., & Kung, H. (2015). An Integrated Environmental Perspective on Software as a Service Adoption in Manufacturing and Retail Firm. *Journal of Information Technology*, 1–12.
Leung, S.-O. (2011). A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales. *Journal of Social Service Research*, *37*(4), 412–421.
Liang, H., Wang, N., & Xue, Y. (2022). Juggling Information Technology (IT) Exploration and Exploitation: A Proportional Balance View of IT Ambidexterity. *Information Systems Research*.
Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*.
Lin, J., Carter, L., Liu, D., Lin, J., Carter, L., & Liu, D. (2021). Privacy Concerns and Digital Government: Exploring Citizen Willingness to Adopt the Covidsafe App. *European Journal of Information Systems*, *30*(4), 389–402.
Lund Research Ltd. (2012). *Constructs in Quantitative Research*. https://dissertation.laerd.com/constructs-in-quantitative-research.php#:~:text=Instead%2C we use variables to operationalize %28i.e.%2C measure%29,the same %28e.g.%2C the construct and variable%2C sex%29.
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), 293–334.

Maier, C., Laumer, S., Tarafdar, M., Mattke, J., Reis, L., & Weitzel, T. (2021). Challenge and Hindrance IS Use Stressors and Appraisals: Explaining Contrarian Associations in Post-Acceptance IS Use Behavior. *Journal of the Association for Information Systems*, *22*(6), 1590–1624.

Maier, C., Thatcher, J. B., Grover, V., & Dwivedi, Y. K. (2023). Cross-Sectional Research: A Critical Perspective, Use Cases, and Recommendations for Is Research. *International Journal of Information Management*, *70*.

Matell, M. S., & Jacoby, J. (1971). Is There an Optimal Number of Alternatives for Likert Scale Items? Study I: Reliability and Validity. *Educational and Psychological Measurement*, *31*, 657–674.

Mayring, P. (2000). Qualitative Content Analysis. *Forum: Qualitative Social Research*, *1*(2).

Mazaheri, E., Lagzian, M., & Hemmat, Z. (2020). Research Directions in Information Systems Field, Current Status and Future Trends: A Literature Analysis of AIS Basket of Top Journals. *Australasian Journal of Information Systems*, *24*, 1–25.

Melnyk, S. A., Page, T. J., Wu, S. J., & Burns, L. A. (2012). Would You Mind Completing This Survey: Assessing the State of Survey Research in Supply Chain Management. *Journal of Purchasing and Supply Management*, *18*(1), 35–45.

Moore, G. C., & Benbasat, I. (1991). Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research*, *2*(3), 192–222.

Moorman, C. (2008). Cross-Sectional Versus Longitudinal Survey Research: Concepts, Findings, and Guidelines. *Journal of Marketing Research*, *45*, 261–279.

Myers, M. D., & Newman, M. (2007). The Qualitative Interview in IS Research: Examining the Craft. *Information and Organization*, *17*(1), 2–26.

Newsted, P. R., Huff, S. L., & Munro, M. C. (1998). Survey Instruments in Information Systems. *MIS Quarterly*, *22*(4), 553–554.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A Method for Taxonomy Development and Its Application in Information Systems. *European Journal of Information Systems*, *22*(3), 336–359.

O'Nyumba, T., Wilson, K., Derrick, C. J., & Mukherjee, N. (2018). The Use of Focus Group Discussion Methodology: Insights From Two Decades of Application in Conservation. *Methods in Ecology and Evolution*, *9*(1), 20–32.

Okoli, C., & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Sprouts: Working Papers on Information Systems*, *10*(26).

Paas, L. J., & Morren, M. (2018). Please Do Not Answer if You Are Reading This: Respondent Attention in Online Panels. *Marketing Letters*, *29*(1), 13–21.

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.

Peter, J. P. (1981). Construct Validity: A Review of Basic Issues and Marketing Practices. *Journal of Marketing Research*, *18*(2), 133.

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

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.

Polites, G. L., Roberts, N., & Thatcher, J. (2012). Conceptualizing Models Using Multidimensional Constructs: A Review and Guidelines for Their Use. *European Journal of Information Systems*, *21*(1), 22–48.

Rai, A., Tang, X., Yin, Z., & Du, S. (2022). Gaining Customer Loyalty with Tracking Information Quality in B2B Logistics. *Journal of Management Information Systems*, *39*(2), 307–335.

Recker, J. (2021). *Scientific Research in Information Systems: A Beginner's Guide* (2nd ed.). Springer Nature Switzerland AG.

Riemenschneider, C. K., & Armstrong, D. J. (2021). The Development of the Perceived Distinctiveness Antecedent Of Information Systems Professional Identity. *MIS Quarterly*, *45*(3), 1149–1186.

Sande, J. B., & Ghosh, M. (2018). Endogeneity in Survey Research. *International Journal of Research in Marketing*, *35*(2), 185–204.

Schmitz, K., Storey, V. C., Schmitz, K., & Robinson, J. M. (2020). Empirical Test Guidelines for Content Validity: Wash, Rinse, and Repeat until Clean. *Communications of the Association for Information Systems*, *47*, 787–850.

Schryen, G. (2015). Writing Qualitative IS Literature Reviews—Guidelines for Synthesis, Interpretation, and Guidance of Research. *Communications of the Association for Information Systems*, *37*(12), 286–325.

Serenko, A., & Turel, O. (2021). Why Are Women Underrepresented in the American IT Industry? The Role of Explicit and Implicit Gender Identities. *Journal of the Association for Information Systems*, *22*(1), 41–66.

Seymour, M., Yuan, L. I., Dennis, A. R., & Riemer, K. (2021). Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments. *Journal of the Association for Information Systems*, *22*(3), 591–617.

Sharma, P. N., Daniel, S. L., Chung, T. R., & Grover, V. (2022). A Motivation-Hygiene Model of Open Source Software Code Contribution and Growth. *Journal of the Association for Information Systems*, *23*(1), 165–195.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *MIS Quarterly*, *20*(2), 167–195.

Srivastava, S. C., Chandra, S., & Shirish, A. (2015). Technostress Creators and Job Outcomes: Theorising the Moderating Influence Of Personality Traits. *Information Systems Journal*, *25*(4), 355–401.

Steelman, Z. R., & Hammer, B. I. (2014). Data Collection in the Digital Age: Innovative Alternatives to Student Samples. *MIS Quarterly*, *38*(2), 355–378.

Straub, D., Gefen, D., & Recker, J. (2022). *Quantitative Research in Information Systems*. Association for Information Systems (AISWorld) Section on IS Research, Methods, and Theories, Last Updated March 25, 2022. http://www.janrecker.com/quantitative-research-in-information-systems/

Tremblay, M. C., Hevner, A. R., & Berndt, D. J. (2010). Focus Groups for Artifact Refinement and Evaluation in Design Research. *Communications of the Association for Information Systems*, *26*.

Trinkle, B. S., Warkentin, M., Malimage, K., & Raddatz, N. (2021). High-Risk Deviant Decisions: Does Neutralization Still Play a Role? *Journal of the Association for Information Systems*, *22*(3), 797–826.

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'Theory' Back Into Grounded Theory: Guidelines for Grounded Theory Studies in Information Systems. *Information Systems Journal*, *20*(4), 357–381.

Venable, J., & Baskerville, R. (2012). Eating Our Own Cooking: Toward a More Rigorous Design Science of Research Methods. *Electronic Journal of Business Research Methods*, *10*(2), 141–153.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative–Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, *37*(1), 21–54.

VERBI Software. (2021). *MAXQDA 2022*. maxqda.com

vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature. *ECIS 2009 Proceedings*.

Voorhees, C. M., Brady, M. K., Calantone, R., Ramirez, E., & Brady, M. K. (2016). Discriminant Validity Testing in Marketing: An Analysis, Causes for Concern, and Proposed Remedies. *Journal of the Academy of Markting Science*, *44*, 119–134.

Wang, J., Li, Y., & Rao, h. R. (2016). Overconfidence in Phishing Email Detection. *Journal of the Association for Information Systems*, *17*(11), 759–783.

Wang, L., Sun, Y., & Luo, X. R. (2022). Game Affordance, Gamer Orientation, and In-Game Purchases: A Hedonic-Instrumental Framework. *Information Systems Journal*, *32*(6), 1097–1125.

Wang, X., & Ji, X. (2020). Sample Size Estimation in Clinical Research From Randomized Controlled Trials to Observational Studies. *Chest*, *158*(1), S12–S20.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2).

Wiener, M., Cram, W., & Benlian, A. (2021). Algorithmic Control and Gig Workers: A Legitimacy Perspective of Uber Drivers. *European Journal of Information Systems*, 1–23.

Wilson, V., Srite, M., & Loiacono, E. (2021). The Effects of Item Ordering on Reproducibility in Information Systems Online Survey Research. *Communications of the Association for Information Systems*, *49*, 760–799.

Zhou, J., Fang, Y., & Grover, V. (2022). Managing Collective Enterprise Information Systems Compliance: A Social and Performance Management Context Perspective. *MIS Quarterly*, *46*(1), 71–99.