Toward Artifact Sampling in IS Design Research

Regular Paper

Roman Lukyanenko
University of Saskatchewan
lukyanenko@edwards.usask.ca

Binny M. Samuel
University of Cincinnati
samuelby@ucmail.uc.edu

Joerg Evermann
Memorial University of Newfoundland
jevermann@mun.ca

Jeffrey Parsons
Memorial University of Newfoundland
jeffreyp@mun.ca

Abstract

Research in natural and social sciences increasingly uses design features of information technologies (IT) as antecedents or outcomes in theoretical models. This practice is also prolific in the Information Systems discipline. The ongoing theorizing of IT leads to a novel methodological challenge termed instantiation validity. In this paper we take a first step to create these guidelines and contribute to research on IV by proposing and advocating the methodological practice of artifact sampling whereby multiple artifacts are (randomly) sampled from the total instantiation space (i.e., the population of all possible artifacts). Artifact sampling extends the prevailing practice of employing multiple research subjects or survey respondents routinely used in social sciences into IT artifact design space. Artifact sampling is an important methodological practice that stands to increase rigor in research dealing with software artifacts, including in the IS discipline. As it is currently not being adequately undertaken in IS research, many studies may result in biased or unjustified conclusions.

Keywords: Instantiation Validity, IS Design Research, Artifact Sampling, Artifact Generalizability, Design Science Research

1. Introduction

As information technology (IT) is central to many human activities, research in natural and social sciences includes design features of IT as antecedents or outcomes in theoretical models. This practice is also prolific in the Information Systems (IS) discipline. In this context, justifying the inclusion of particular design choices in an IT artifact is a key challenge facing researchers. We focus on the case in which a researcher is interested in evaluating a theory in which information technology is conceptualized as an independent variable or when researchers seek to evaluate a design science research theory (or design principle). We term studies that use IT artifacts as part of theory testing or to evaluate design principle ***information systems design research (ISDR)***[1].

Consider several example of IS design research. Behavioral IS researchers may posit that creating personalized recommender systems leads to the adoption of these systems by online users due to the greater propensity of personalized technologies to engender trust with users (e.g., see Aksoy et al. 2006; Benbasat and Wang 2005; Komiak and Benbasat 2004). Likewise, design science researchers in IS may wish to evaluate a novel design proposition that relaxing protocols to collect information from crowds (through instance-based as opposed to the traditional, class-based conceptual and data modeling) leads to increased accuracy and completeness of data provided by the non-expert crowds (Lukyanenko, Parsons, et al. 2014a, 2014b; Lukyanenko and Parsons 2015).

In order to evaluate these general ideas, researchers from both strands engage in ***design work*** to select or build one or more IT artifacts that correspond to desired levels of personalization (e.g., Komiak and Benbasat 2006) or that implement the proposed design principles (e.g., Lukyanenko et al. 2016). These artifacts could then be used by research participants, who report their experiences to the researchers. These responses are then used to test the underlying theory of personalized technologies or design theory of modeling and information quality. In both scenarios, the research findings and conclusions depend critically on the design decisions taken when selecting or creating the IT artifacts (i.e., during the design of the artifact itself).

The ongoing theorizing of IT has resulted in methodological challenges (Benbasat 1989; Boudreau et al. 2001; Burton-Jones et al. 2009; Orlikowski and Iacono 2001; Parsons and Cole 2005; Straub et al. 2004). According to Lukyanenko et al. (2014), when instantiating a particular theoretical construct there are virtually unlimited ways to operationalize (i.e., design) the feature in the corresponding IT artifact, but no clear guidance for choosing the most appropriate one. Further, while a researcher may be only interested in one particular construct (e.g., personalization) the artifact that instantiates that construct often has to include a variety of features (e.g., navigation/help buttons) to provide basic functionality and usability. These features are not chosen based on instantiating the construct of interest, but may interact with this construct in unpredictable ways, potentially affecting results and diminishing internal validity. These concerns have resulted in a proposal for a new kind of research validity (Cook and Campbell 1979) – ***instantiation validity*** (IV) – defined as the extent to which inferences and conclusions are justified and warranted from investigations of IT artifacts as instantiations of theoretical constructs or design principles (Lukyanenko, Evermann, et al. 2014).

While instantiation validity is an emerging concern, little practical methodological guidance for establishing and demonstrating IV exists (for review, see Lukyanenko et al. 2015). We take a first step to create these guidelines and contribute to research on IV by proposing and advocating the

[1] IS design research, as conceptualized in Lukyanenko et al. (2014), is somewhat different from information systems design science research (DSR). Unlike a typical DSR, IS design research is narrowly focused on the creation of artifacts for the purpose of theory testing (as in IS behavioral research) or instantiation of an abstract idea (as in some DSR, see (Purao 2013)), and, therefore, does not focus on the problem-solving activities of DSR (Gregor and Hevner 2013; Hevner and Chatterjee 2010; Venable 2015). Furthermore, IT artifacts built in a behavioral design study need not be of innovative nature (and could reuse existing technology). At the same time, both DSR and IS design research, are concerned with methodological rigor when developing and evaluating IT artifacts (Hevner and Chatterjee 2010; Peffers et al. 2007; Prat et al. 2015; Tremblay et al. 2010; Venable et al. 2012).

methodological practice of ***artifact sampling*** whereby multiple artifacts are (randomly) sampled from the total instantiation space (i.e., the population of all possible artifacts).

2. Roots of Artifact Sampling in Psychology and Sampling Theory

The major threats to instantiation validity include the complexity of a typical software artifact, potentially creating confounds arising from the interaction of the focal features with non-focal features of the artifact, as well as a vast and potentially unbounded space of possible implementations (Lukyanenko, Evermann, et al. 2014). We propose that one way to address these threats is by increasing the number of the artifacts, analogous to the way researchers routinely increase the number of human participants to reduce sampling error or increase the number of questionnaire items to improve reliability. Such an approach is proposed as methodological guidance during the design process.

Sampling theory underlies much of scientific experimental work (Lohr 2009). Fundamental to the theory is the principle that one may generalize the results of observations only to those subjects or objects that have been sampled (Hammond and Stewart 2001). As early as 1940s, however, researchers pointed out a peculiar "double standard" (Brunswik 1955; Hammond and Stewart 2001). Researchers were quite eager to apply sampling theory to *subjects* (e.g., human participants, survey respondents), but almost never extended this principle to research *objects* (i.e., experimental stimuli) (Brunswik 1943). Even more concerning, Brunswik argued, is that over time, researchers developed a variety of systematic approaches to increase rigor in subject sampling, including statistical methods to determine sample sizes, estimate errors and biases and draw statistical inferences. Thus, seeking large sample sizes offers an ability to eliminate potentially idiosyncratic effects of differences among individual subjects (Cook and Campbell 1979; Highhouse 2009). The theoretical premise is that the differences are assumed to be independent of: 1) any treatment effect; 2) each other; 3) and across subjects. Therefore, the subject differences "cancel each other out" in a sufficiently large sample. In the meantime, little attention has been paid to research objects. As early as in 1943, Brunswik (1943) introduced the notion of *representative designs* which argues that sampling theory *equally concerns* subjects and objects of research. Yet, the recognition of this idea has been slow. Among key objections to Brunswik's (1943) argument was the effort involved in sampling objects – an argument that persists (see, e.g., Lindzey et al. 1998).

Recently, the idea of having multiple objects within treatment and control conditions, has been gaining acceptance in psychology. Echoing the instantiation validity concerns described earlier, psychologists argue and show experimentally that it is generally impossible to construct ecologically valid objects such that every feature is accounted for theoretically, and that it is difficult for researchers to adequately (i.e., fully) represent and generalize to a population of objects from a single object (Bonge et al. 1992; Dhami et al. 2004; Fontenelle et al. 1985; Hammond and Stewart 2001; Highhouse 2009; Kelley 1992; Snodgrass and Vanderwart 1980; Wells and Windschitl 1999). This appears to be the case both for complex objects (e.g., humans – often used to instantiate independent variables in social psychology, see (Wells and Windschitl 1999)) as well as simpler objects (e.g., line drawings, see (Snodgrass and Vanderwart 1980)) commonly used in cognitive psychology. Even when the objects are quite simple (i.e., have few features and potential interactions between them), Fontenelle et al. (1985) conclude: "when it is the intention of an experimenter to generalize results beyond the particular sample of objects employed, the statistical treatment of objects as a fixed effect is generally inappropriate. Thus, unless a researcher is willing to limit the generalizability of his or her findings *severely*, the effect of stimulus sampling must be considered both in the design of the experiment and in the analysis of the results." (p. 106, emphasis added).

While the benefits of involving multiple subjects in experiments and surveys have been widely recognized, the second part of the original representative design notion that suggested to do the same for objects have been neglected in experimental research. Wells and Windschitl (1999, p. 1115) consider this neglect "a serious problem that plagues a surprising number of experiments," casting doubts on the validity of conclusions drawn from such studies. To increase the validity of experimental studies, more and more researchers call for ***stimuli sampling*** – selecting objects at random from the theoretical feature space (Fontenelle et al. 1985; Wells and Windschitl 1999).

Sampling from a design space also occurs in the construction and validation of surveys instruments for psychometric research in IS. Straub (Straub 1989, p. 150), citing Cronbach (Cronbach 1971), notes that "an instrument valid in content is one that has drawn representative questions from a universal pool". Similarly, we propose that an artifact that is valid in content with respect to a construct is one that has features drawn in a representative way from a universal pool (of possible features that might instantiate the construct in an artifact). Straub further suggests that "a content-valid instrument is difficult to create ... because the universe of possible content is virtually infinite" (page 150). Again, referring to Cronbach (Cronbach 1971), Straub recommends an expert to evaluate the instruments. This recommendation for establishing content validity for survey instruments with the help of expert assessment has been adopted in the recommendation of focus groups (Tremblay et al. 2010) for instantiation validity by Lukyanenko et al. (2015) .

We extend this suggestion of sampling object stimuli (experimental or questionnaire items) to the sampling of artifacts and features in IS design research. As mentioned earlier, the problems of IV, while present in other disciplines, are particularly important for studies involving IT. Unlike simple drawings, silhouettes, stick figures common in psychology (e.g., Beeck and Wagemans 2001; Rosch et al. 1976), IT are considerably more complex. The patterns of interaction with IT are constantly evolving, further confounding efforts to detect extraneous interferences.

3. Treating Artifacts as Research Subjects: Toward an Artifact Sampling Method

Motivated by the methodological suggestions and arguments in social sciences, here we consider some options for ***artifact sampling.***

Artifact sampling extends the concept of stimulus sampling from experimental psychology and scale reliability from survey research to research involving software artifacts. Effectively we propose an analogy to treat artifacts as studies routinely treat research subjects to deal with vast individual variability and large total population.

Artifact sampling entails selecting multiple artifacts from the space of valid design possibilities. Software artifacts are intended to instantiate, through certain features, a particular level of one or more theoretical constructs, for example a high degree of personalization. Given the typically very large, if not infinite, design space of these features, a sampling from this design space produces a set of artifacts that are representative of the desired theoretical construct level, e.g. high personalization.

Instrument validation in survey research establishes construct validity by answering the question whether “instruments show stability across methodologies”. In other words, construct validity “asks whether the measures chosen are ... merely artifacts of the [measurement] methodology itself” (Straub 1989, p. 150). The immediate parallel in instantiation validity is the question whether the instantiation is biased by its construction methodology (Lukyanenko et al., 2014). To answer this question, different artifacts may be sampled from different construction methods (e.g. web-based, mobile app), interface paradigms (e.g. mouse, touch, VR), or application domains (e.g. financial services, social networking, e-commerce) to enable identifying the influence of any of these factors on the artifact as necessary to ensure the external validity claimed by the researchers.

For example, a researcher looking to understand the ability of IT to convey emotions (a theoretical construct) may propose a multitude of possible designs corresponding to specific ways this construct may be implemented in line with the proposed construct definition. This first entails constructing an ***instantiation space*** by closely examining the theory and deriving from it a conceptual space of valid implementations. The process of identifying a theoretical space and deriving multiple objects that instantiate it is becoming better understood in psychology, as it develops stimuli libraries (e.g., Alario and Ferrand 1999; Berman et al. 1989). From this work, it is evident that this process requires deeper theoretical rigor, as it involves deep and thorough understanding of what makes an implementation a valid instance of the construct (Snodgrass and Vanderwart 1980). Here, design science research in IS, in particular, stands to inform artifact sampling, as it has a tradition of working with artifacts at concrete and abstract levels (see, e.g., Arazy and Kopak 2011; Chandra et al. 2015; Chandra Kruse et al. 2016; Gregor et al. 2013; Lukyanenko et al. 2015; Prat et al. 2015; Purao 2013).

These implementations need not consider every possible way to implement the construct (now and in the future) but, as argued by Wells and Windschitl (1999, p. 1115), should be representative enough and contain enough variation to capture as many possible confounds as feasible for the project (see, Baskerville and Pries-Heje 2014; Lukyanenko and Parsons 2013). Constructing an instantiation space therefore requires both deep understanding of the construct and of the design possibilities (Arazy et al. 2010). Returning to the “ability of IT to convey emotions” construct, researchers may thus conceive of various ways to implement specific levels of this construct, using different construction methods, different interface elements and in different application domains.

Once the theoretical sample space is established, sampling procedures should be applied to select multiple artifacts, which are then implemented and used for evaluation. The sample size and its selection is naturally constrained by:

- (expected) natural variability of relevant features in the population of artifacts (where greater variability calls for more artifacts);
- expected confounding factors and the difficulty in detecting and controlling (here, more artifacts could be used, at least, partially to assuage concerns about potential confounds);
- desire to draw stronger inferences (which may suggest striving for larger sample sizes, random selection to perform analysis of variance tests over groups of artifacts); and
- pragmatic considerations (e.g., cost, effort of implementation, which may limit the number of artifacts).

One potential suggestion is to echo the sentiments from multi-item measurement for survey scale development that encourages 3-7 items per construct to achieve adequate reliability

(Diamantopoulos et al. 2012). While this guidance is tentative at best, it provides a starting point to compare to a single instantiation.

Similarly, artifacts may be sampled from the instantiation space to represent different but related constructs. In the example of "ability to convey emotions", related constructs might be "media richness" or "visual appeal". As with the multi-trait, multi-method approach in psychometric work, artifacts instantiating the same construct but developed using different construction methods, interface paradigms, or application domains should lead to similar consequences, whereas artifacts representing different constructs, even though using the same construction method, interface paradigm, or application domain, should lead to dissimilar consequences. While these consequences are typically numeric ratings by survey respondents in traditional psychometric work, the parallel in design science and instantiation validity may be similar observed usage patterns or some other measurable consequences of the artifact; it need not be a quantitative response.

Finally, artifact sampling may be pursued in three ways:

- Sampling for artifact diversity;
- Sampling for artifact homogeneity; and
- Sampling for diversity among sets of similar artifacts.

First, researchers can sample for artifact diversity and breadth to cover many points in the design space. The aim here to improve generalizability (i.e. inference to the population) and get an assessment of the heterogeneity of the design space (which will inform any generalizability claims one makes). In the second case, researchers sample the same (or very similar) point in the design space for homogeneity to get a more reliable sample and reliable theoretical claims. Here, the aim is homogeneity of the sample so that minor local variations of the design space "cancel each other out". Finally, researchers may combine the two strategies above to obtain heterogenous set of homogenous sets of samples, which would allow reliable claims about each sampled point and also allow claims to generalizability based on a thorough understanding of the different parts of the design space. Thus, we strive for the latter to the extent possible.

4. Conclusions and Future Work

We propose a novel methodological concept – ***artifact sampling*** – intended to increase the instantiation validity of studies involving software artifacts. Artifact sampling is an important methodological practice that stands to increase rigor in research dealing with software artifacts, including in the IS discipline. Furthermore, we suggest it is not being adequately undertaken in IS research to date and may result in biased conclusions.

The key contribution of this research is to motivate future work on the method of artifact sampling. To pave the way for future work, we have provided the background for artifact sampling, showing in particular that it has deep methodological foundations in sampling theory, the notion of representative design and is akin to the well-established norms for increasing reliability in psychometric research. The idea of artifact sampling is being increasingly accepted in psychology, where it is known as stimuli sampling. Recently, renewed and stronger arguments in favor of stimuli sampling are being made, new approaches and methods are being proposed and, further underscoring the on-going acceptance of the idea, libraries of stimuli are proliferating at a rapid rate (see references above). This motivated us to consider the implications of these developments for IS that manifested in the artifact sampling proposal.

Admittedly, artifact sampling method calls for more resources and theoretical rigor and we acknowledge such limitations in our early stage research and call for more research to help provide specific guidelines. First, artifact sampling may not be always be useful, just as in some cases a single-item survey scale is sufficient (Diamantopoulos et al. 2012). For example, artifact sampling may not be needed if testing the effect of Facebook (as *the* focal social network site) use if there is no intent to generalize to other social media technologies. Indeed, sampling potential social network artifacts may not be practical or useful in such situations. However, as the space of social networking sites is rapidly expanding (Culnan et al. 2010; Kane et al. 2014; Maddah et al. 2016), testing Facebook alone may not be sufficient if one wishes to make conclusions for the general population of social networks. Likewise, if an artifact has wide acceptance, it may be useful to study its effects without sampling. Artifact sampling is more geared toward nomothetic rather than idiographic research objectives (see, Amrollahi and Lukyanenko 2016; Lukyanenko and Darcy 2016).

Second, guidelines on how to establish the instantiation space are needed to help researchers carefully plan out their instantiation options. The dimensions of the design space should be orthogonal, as much as possible, to ensure that the artifacts that are sampled from this are independent. Third, guidelines to establish the independence of the sampled items as well as the number of items are necessary. Fourth, the development of quantitative or qualitative techniques that allow subjects to evaluate the instantiation validity of objects is necessary.

Clearly, artifact sampling will not apply to cases where the instantiation space is limited and small and where the dimensions of the space cannot be defined independently of each other. However, as argued in (Lukyanenko, Evermann, et al. 2014; Lukyanenko and Parsons 2013), many IS research questions deal with situations where it is unclear how to design an artifact and many (and sometimes potentially an unlimited) number of design choices exist. Indeed, the notion of a potentially vast space of possible operationalizations is recognized in other disciplines (Wells and Windschitl 1999), and we believe it should at least be considered in IS research, especially during the process of designing and evaluating artifacts. Importantly, however, this process holds IT-based research to higher standards as it helps to addresses instantiation validity concerns and increase the confidence in the conclusions of IS design studies. It also opens a variety of novel and intriguing methodological possibilities, promising better science and advancing IT design knowledge.

We acknowledge that the notion of artifact sampling for instantiation validity might be met with its own criticisms. For example, some may argue that design decisions are ultimately guided by theory, and not empirical evaluation (a position we also hold, but we suggest that often it is difficult to settle on a single correct design). Drawing from the methodological context of scale development research, the choice of whether to drop/add an item is ultimately determined by theoretical reasons, not just the empirical evaluation. However, empirical measurement model techniques do provide recommendations with respect to how valid the measurement of the construct is with the presence/absence of the item. Another criticism may be the notion of a program of study (Lakatos and Musgrave 1970) and/or replication of a design to ultimately find the appropriate operationalization. For example, maybe in the initial operationalization of the design, providing a definitive theoretical justification for design choices is impractical, and, further studies can help refine and confirm the validity of the design choices (e.g., Tremblay et al. 2010). We believe this approach is also sound. Future studies should explore in greater detail when artifact sampling is more effective and epistemically appropriate and when other strategies should be pursued.

In the future, we hope to better understand the process of artifact sampling, develop best practices, address the issue of when this method should be applied and provide specific examples that illustrate application of this concept. We also hope that this paper will motivate further discussions about both the proposed idea of artifact sampling and the broader concerns of instantiation validity.

5. References

Aksoy, L., Bloom, P. N., Lurie, N. H., and Cooil, B. 2006. "Should Recommendation Agents Think Like People?," *Journal of Service Research* (8:4), pp. 297–315.

Alario, F.-X., and Ferrand, L. 1999. "A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition," *Behavior Research Methods, Instruments, & Computers* (31:3), pp. 531–552.

Amrollahi, A., and Lukyanenko, R. 2016. "A Multi-Paradigmatic Approach to Theorising in Design Science Research," in *Pre-ACIS 2016 workshop on "Advances in Qualitative IS Research Methodologies,"* Presented at the Pre-ACIS 2016 Workshop on "Advances in Qualitative IS Research Methodologies," Sydney, Australia.

Arazy, O., and Kopak, R. 2011. "On the measurability of information quality," *Journal of the American Society for Information Science and Technology* (62:1), pp. 89–99.

Arazy, O., Kumar, N., and Shapira, B. 2010. "A theory-driven design framework for social recommender systems," *Journal of the Association for Information Systems* (11:9), pp. 455–490.

Baskerville, R., and Pries-Heje, J. 2014. "Design Theory Projectability," in *Information Systems and Global Assemblages.(Re) Configuring Actors, Artefacts, Organizations*, Springer, pp. 219–232.

Beeck, H. O. de, and Wagemans, J. 2001. "Visual object categorisation at distinct levels of abstraction: A new stimulus set," *Perception* (30:11), pp. 1337–1361.

Benbasat, I. 1989. *Laboratory experiments in information systems studies with a focus on individuals: a critical appraisal* (Vol. 2), Cambridge, MA: Harvard Business School.

Benbasat, I., and Wang, W. 2005. "Trust In and Adoption of Online Recommendation Agents," *Journal of the Association for Information Systems* (6:3), pp. 72–101.

Berman, S., Friedman, D., Hamberger, M., and Snodgrass, J. G. 1989. "Developmental picture norms: Relationships between name agreement, familiarity, and visual complexity for child and adult ratings of two sets of line drawings," *Behavior Research Methods, Instruments, & Computers* (21:3), pp. 371–382.

Bonge, D. R., Schuldt, W. J., and Harper, Y. Y. 1992. "The experimenter-as-fixed-effect fallacy," *The Journal of psychology* (126:5), pp. 477–486.

Boudreau, M.-C., Gefen, D., and Straub, D. W. 2001. "Validation in Information Systems Research: A State-of-the-Art Assessment," *MIS Quarterly* (25:1), pp. 1–16.

Brunswik, E. 1943. "Organismic achievement and environmental probability.," *Psychological review* (50:3), p. 255.

Brunswik, E. 1955. "Representative design and probabilistic theory in a functional psychology.," *Psychological review* (62:3), p. 193.

Burton-Jones, A., Wand, Y., and Weber, R. 2009. "Guidelines for Empirical Evaluations of Conceptual Modeling Grammars," *Journal of the Association for Information Systems* (10:6), pp. 495–532.

Chandra, L., Seidel, S., and Gregor, S. 2015. "Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions," *Hawaii International Conference on System Sciences*, pp. 4039–4047.

Chandra Kruse, L., Seidel, S., and Purao, S. 2016. "Making Use of Design Principles," Presented at the DESRIST 2016, LNCS 9661, pp. 37–51.

Cook, T. D., and Campbell, D. T. 1979. *Quasi-experimentation : design & analysis issues for field settings*, Chicago: Rand McNally College Pub. Co.

Cronbach, L. J. 1971. "Test validation," in *Educational measurement*R. Thorndike (ed.), pp. 443–507.

Culnan, M. J., McHugh, P. J., and Zubillaga, J. I. 2010. "How large U.S. companies can use Twitter and other social media to gain business value.," *MIS Quarterly Executive* (9:4), p. 243.

Dhami, M. K., Hertwig, R., and Hoffrage, U. 2004. "The role of representative design in an ecological approach to cognition.," *Psychological bulletin* (130:6), p. 959.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., and Kaiser, S. 2012. "Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective," *Journal of the Academy of Marketing Science* (40:3), pp. 434–449 (doi: 10.1007/s11747-011-0300-3).

Fontenelle, G. A., Phillips, A. P., and Lane, D. M. 1985. "Generalizing across stimuli as well as subjects: A neglected aspect of external validity.," *Journal of Applied Psychology* (70:1), p. 101.

Gregor, S., and Hevner, A. R. 2013. "Positioning and presenting design science research for maximum impact," *MIS Quarterly* (37:2), pp. 337–355.

Gregor, S., Müller, O., and Seidel, S. 2013. "Reflection, Abstraction And Theorizing In Design And Development Research.," Presented at the ECIS, p. 74.

Hammond, K. R., and Stewart, T. R. 2001. *The essential Brunswik: Beginnings, explications, applications.*, Oxford University Press.

Hevner, A., and Chatterjee, S. 2010. *Design Research in Information Systems: Theory and Practice* (Vol. 22), Springer.

Highhouse, S. 2009. "Designing experiments that generalize," *Organizational Research Methods* (12:3), pp. 554–566.

Kane, G. C., Alavi, M., Labianca, G. J., and Borgatti, S. 2014. "What's different about social media networks? A framework and research agenda," *MIS Quarterly* (38:1), pp. 274–304.

Kelley, L. 1992. *Issues, theory, and research in industrial/organizational psychology* (Vol. 82), Elsevier.

Komiak, S. X., and Benbasat, I. 2004. "Understanding Customer Trust in Agent-Mediated Electronic Commerce, Web-Mediated Electronic Commerce, and Traditional Commerce," *Information Technology and Management* (5:1), pp. 181–207.

Komiak, S. Y. X., and Benbasat, I. 2006. "The effects of personalization and familiarity on trust and adoption of recommendation agents.," *MIS Quarterly* (30:4), pp. 941–960.

Lakatos, I., and Musgrave, A. 1970. *Criticism and the Growth of Knowledge*, Cambridge England ; New York: Cambridge University Press.

Lindzey, G., Gilbert, D., and Fiske, S. T. 1998. *The handbook of social psychology*, Oxford University Press.

Lohr, S. L. 2009. *Sampling: Design and Analysis*Advanced (Cengage Learning), Cengage Learning (available at https://books.google.com/books?id=aSXKXbyNlMQC).

Lukyanenko, R., and Darcy, D. 2016. "On Systematicity: Expanding the Diversity of Design Science Research Contributions," in *DESRIST 2016, Tackling Society's Grand Challenges with Design Science*, St. John's, NL Canada, pp. 1–7.

Lukyanenko, R., Evermann, J., and Parsons, J. 2014. "Instantiation Validity in IS Design Research," in *DESRIST 2014, LNCS 8463*, Springer, pp. 321–328.

Lukyanenko, R., Evermann, J., and Parsons, J. 2015. "Guidelines for Establishing Instantiation Validity in IT Artifacts: A Survey of IS Research," in *DESRIST 2015, LNCS 9073*, Berlin / Heidelberg: Springer.

Lukyanenko, R., and Parsons, J. 2013. "Reconciling theories with design choices in design science research," in *DESRIST 2013, LNCS 7939*Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pp. 165–180.

Lukyanenko, R., and Parsons, J. 2015. "Principles for Modeling User-Generated Content," in *Conceptual Modeling*, Springer, pp. 432–440.

Lukyanenko, R., Parsons, J., and Wiersma, Y. 2014a. “The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-generated Content,” *Information Systems Research* (25:4), pp. 669–689.

Lukyanenko, R., Parsons, J., and Wiersma, Y. 2014b. “The Impact of Conceptual Modeling on Dataset Completeness: A Field Experiment,” Presented at the International Conference on Information Systems, pp. 1–18.

Lukyanenko, R., Parsons, J., Wiersma, Y., Sieber, R., and Maddah, M. 2016. “Participatory Design for User-generated Content: Understanding the challenges and moving forward,” *Scandinavian Journal of Information Systems* (28:1), pp. 37–70.

Maddah, M., EsmaeilZadeh, P., Cousins, K., and Lukyanenko, R. 2016. “Why Use Social Networks? Toward a Comprehensive Framework,” in *Twenty-second Americas Conference on Information Systems*, Presented at the Americas Conference on Information Systems, San Diego, CA, USA.

Orlikowski, W. J., and Iacono, C. S. 2001. “Research commentary: Desperately seeking the IT in IT research—a call to theorizing the IT artifact,” *Information systems research* (12:2), pp. 121–134.

Parsons, J., and Cole, L. 2005. “What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques,” *Data & Knowledge Engineering* (55:3), pp. 327–342.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. “A design science research methodology for information systems research,” *Journal of management information systems* (24:3), pp. 45–77.

Prat, N., Comyn-Wattiau, I., and Akoka, J. 2015. “A Taxonomy of Evaluation Methods for Information Systems Artifacts,” *Journal of Management Information Systems* (32:3), pp. 229–267.

Purao, S. 2013. “Truth or dare: The ontology question in design science research,” *Journal of Database Management (JDM)* (24:3), pp. 51–66.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyesbraem, P. 1976. “Basic Objects in Natural Categories,” *Cognitive Psychology* (8:3), pp. 382–439.

Snodgrass, J. G., and Vanderwart, M. 1980. “A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity,” *Journal of Experimental Psychology: Human Learning and Memory* (6:2), pp. 174–215.

Straub, D., Boudreau, M.-C., and Gefen, D. 2004. “Validation guidelines for IS positivist research,” *The Communications of the Association for Information Systems* (13:1), pp. 1–63.

Straub, D. W. 1989. “Validating Instruments in MIS Research,” *MIS Quarterly* (13:2), pp. 147–169.

Tremblay, M. C., Hevner, A. R., and Berndt, D. J. 2010. “Focus Groups for Artifact Refinement and Evaluation in Design Research.,” *Communications of the Association for Information Systems* (26), pp. 599–618.

Venable, J., Pries-Heje, J., and Baskerville, R. 2012. “A comprehensive framework for evaluation in design science research,” in *DESRIST 2012, LNCS 7286*, Springer, pp. 423–438.

Venable, J. R. 2015. “Five and Ten Years on: Have DSR Standards Changed?,” in *DESRIST 2015, LNCS 9073*, Springer, pp. 264–279.

Wells, G. L., and Windschitl, P. D. 1999. “Stimulus sampling and social psychological experimentation,” *Personality and Social Psychology Bulletin* (25:9), pp. 1115–1125.