Guidelines for Establishing Instantiation Validity in IT Artifacts: A Survey of IS Research

Roman Lukyanenko[1], Joerg Evermann[2], Jeffrey Parsons[2]

[1] College of Business, Florida International University, Miami, FL USA
roman.lukyanenko@fiu.edu
[2] Faculty of Business Administration, Memorial University, St. John's, NL Canada
{jevermann@mun.ca, jeffreyp}@mun.ca

Abstract. The centrality of information technology (IT) artifacts in Information Systems (IS) research makes it important to understand the relationship between artifacts and the theoretical constructs they purport to instantiate. Despite the central role of the IT artifact in IS research, there are no generally accepted principles for establishing *instantiation validity* – the extent to which an artifact is a valid instantiation of a theoretical construct or a manifestation of a design principle. We survey relevant knowledge in IS and identify potential guidelines that may address threats to instantiation validity. The guidelines are intended for researchers and reviewers when using IT artifacts in theory testing and when evaluating design science artifacts.

Keywords: Instantiation Validity, IT Artifact, Design Science Research, Methodology.

1 Introduction

Information Systems (IS) research routinely conceptualizes properties of IT artifacts as theoretical constructs that impact human behavior of interest, or as manifestations of design principles intended to achieve some outcome. Properties of IT artifacts are central to design science research (DSR) in IS, which develops and evaluates constructs, models, methods, implementations, and design theories.

A common practice in IS research is manipulating features of an IT artifact to evaluate theoretical models. For example, Komiak and Benbasat [1] investigate how "personalization" and "familiarity" affect IT adoption. They selected two existing software systems (recommendation agents) assumed to correspond to different levels of perceived personalization and familiarity. They conducted an experiment demonstrating that different levels of perceived personalization and familiarity engender different levels of intention to adopt. According to Lukyanenko et al. [2], the validity of this conclusion (i.e., that personalization of an IS leads to its increased adoption) "depends critically on whether the chosen artifacts faithfully instantiated the underlying theoretical construct of personalization and levels thereof" (p. 322). Lukyanenko et al. introduce the notion of ***instantiation validity*** (IV) to denote "validity of IT artifacts as instantiations of theoretical constructs" (ibid).

In a similar vein, design science research is concerned with the construction of IT artifacts that manifest certain design principles intended to solve particular problems.

adfa, p. 1, 2015.
© Springer-Verlag Berlin Heidelberg 2015

In that case, instantiation validity is the extent to which the artifacts are consistent with the design principles. For conciseness, in this paper we intend the phrase "theoretical constructs" to include design principles where appropriate.

Instantiation validity is fundamentally concerned with the relationship between ***abstract*** theoretical constructs and IT artifacts - ***concrete*** software systems that are intended to instantiate (levels of) one or more constructs or design principles. IV takes on aspects of internal and construct validity, but the uniqueness of IV that arises from distinctive properties of IT artifacts and characteristics of the IS domain (e.g., relentless technological progress) limits the useful guidance from traditional research on validity [3, 4].

While Lukyanenko et al. [2] discuss IV in the context of IS theory testing, we believe this concept also applies to other types of knowledge contribution in IS research. Instantiation validity plays a role in evaluating design knowledge when this knowledge is used to instantiate properties of an IT artifact. Indeed, Venable et al. [5] note that, in evaluating an artifact for its utility, a researcher also evaluates the design theory that the artifact is based on. Rossi and Sein [6] suggest that evaluation should include assessing the match between an "abstract idea" and the artifact. Venable at al. [5] provide a comprehensive framework for artifact evaluation. Hevner and Chatterjee [7] classify evaluation into analytical modeling, simulation and measurement-based strategies. Of these, simulation (writing software code to mimic behavior of the proposed system) and measurement (e.g., experimentation using human participants and real systems) may result in a "medium shift" when abstract design principles are transformed into concrete forms, [2] thereby creating IV challenges.[1] Design science researchers increasingly call for a more transparent DSR process. Specifically, an evaluation stage has been suggested not only after the artifact is built, but also before and during development [8, 9]. IV can be conceptualized at the core of the design process as it ensures that the design process is transparent and justified.

Further, IV is a challenge for practitioners looking to build systems based on design knowledge. While design knowledge is generated in some specific context, once it is finalized or deemed sufficiently complete [10], researchers strive to generalize beyond the specific context of creation to inform development in other settings. The final product of design theorizing aims to solve some class of real-world problems [11, 12] by virtue of causal impact of artifact properties on features of the environment [13]. Chandra et al. [14] note that IV answers the question of whether a real-world artifact that adopts some design theory "indeed proffers the action described by the design principle" (p. 4046). Similar to theory testing, failure to build an artifact sufficiently similar to the one envisioned by the researcher may result in failure to solve the problem for which the artifact is built. Unless practitioners are aware of IV as a threat, this failure could undermine the **perceived** credibility of design knowledge.

In summary, the need to establish or demonstrate IV arises during: ***theory testing***, when IS artifacts are used to manipulate theoretical constructs [2]; ***evaluation*** of a

[1] Note, analytical modeling may also involve a medium shift if the original design knowledge is expressed in natural language and is transformed into a symbolic representation.

design science artifact, if the evaluation relies on a concrete instantiation [5, 14]; and the ***application*** of IS design knowledge by the practitioner community.

Despite the importance of IV, there are no generally accepted principles for establishing and demonstrating it. Lukyanenko et al. [2] call for establishing such principles. This paper attempts to answer this call. We synthesize relevant knowledge accumulated in IS to address the five threats to IV proposed in [2]. A review of IS reveals potentially useful recommendations that can guide researchers, reviewers and practitioners in theory testing, design science theorizing and evaluation and real software development.

2 The Instantiation Validity Problem in IS Research

IS researchers have long emphasized methodological rigor, in which establishing and demonstrating validity plays an important role. Commonly, rigor is demonstrated with respect to validity notions in social sciences (e.g., internal, construct, content, predictive) [3, 15], without specifically considering the validity of the artifact itself [2]. In particular, a seminal paper on the "state-of-the-art" of validation in IS by Boudreau et al. [3] conceptualized validity as internal, content, and construct validity. In the context of DSR rigor, Hevner at al. [11] likewise do not discuss the validity of an IT artifact as embodiment of a construct.

A major approach to validation in IS is to demonstrate validity post hoc using manipulation checks and statistical techniques. The idea of a manipulation check is to ensure that research subjects have perceived the intended manipulation of a theoretical construct [4], i.e. to assess the extent to which they have received the intended experimental treatment. For Boudreau et al. [3] manipulation checks are "critical tests of instrumentation" [i.e., the artifact] (p. 5). Unfortunately, however, manipulation checks can only be used once the system is built – they offer no guidance for developing valid instantiations. This is particularly problematic given the cost of artifact development. Furthermore, when the construct affects "hidden" features, such as an underlying algorithm, a manipulation check based on perceptions may not be possible. Manipulation checks may also "frame" the problem by providing research subjects with cues that lead them to answers they may not otherwise provide.

As using real or realistic software systems is costly (an IV threat [2]), IS research has proposed a number of potentially applicable strategies for pre-development validation. Notably, Benbasat [16] discusses the properties of experimental stimuli under the label "research design" and characterizes a design as faulty when a stimulus does not clearly separate the focal theoretical construct from others. Benbasat [ibid.] suggests that "the major point … is to first determine precisely on what basis the stimulus materials are to be different" (p. 42) and that "once this is known, it becomes easier to determine if equivalency, except for the stimulus in question, was achieved" (p. 42).

While some research distinguishes between evaluating features of an artifact before and after instantiation [e.g., 5, 17], others, including Sein et al. [18], Eriksson et al.[9] and Abraham et al. [19] argue for evaluation conducted during development. We adopt the recommendation of concurrent evaluation as it helps to address the threats to IV posed due to high cost of development and artifact complexity.

Another guideline is to re-use existing measurement items or employ standardized items (e.g., [20, 21]) or standardized stimuli (e.g., [22]). Stimuli standardization is less common in IS, but common in reference disciplines, such as psychology (e.g, [22]). While standardization has been popular in survey and experimental research, IT artifacts are frequently modified in response to rapid technological progress (an IV threat [2]), thereby limiting the extent of artifact standardization in IS research.

Another important concept found in IS research is that of traceability of theoretical constructs from more general to more specific ones. The DSR community has proposed approaches to link theoretical propositions from reference theories (known as justificatory knowledge or kernel theories) to design principles [11], recognizing the need to make a given theory and its constructs more concrete if these are to inform the construction of an artifact. While the focus has been on design principles or survey of reference disciplines (e.g., [23]), rather than the implemented artifact and its relationship with the design principles, the recommendation to provide a transparent link between different levels of abstraction apply to the concept of IV.

The long tradition of experimenting and building IT artifacts by IS researchers has produced a number of valuable strategies for both pre-development and post hoc validation of artifacts and measurement items. IS researchers clearly demonstrate awareness of the potential problems with using complex software systems as scientific instruments but, at the same time, lack widely-agreed criteria for the demonstrating and establishing IV. To develop such guidelines , in the next section we apply the suggestions from previous research to specific validity threats introduced in [2].

3 Addressing Threats to Instantiation Validity

In this section, we use the ***five-threat IV framework*** [2] to develop guidelines for establishing and demonstrating IV based on existing thinking in IS research.

Artifact Cost. The construction of typical IT artifacts is relatively more expensive than survey instruments. As a result, researchers may have the resources to create only a single artifact with limited functionality (e.g., in contrast to multiple measurement items). This limits the ability to control for confounding effects, to demonstrate validity and reliability by comparing multiple implementations, and to test multiple (especially extreme) levels of a construct.

One way of addressing this challenge is to implement features in a flexible, parameterized way. This allows researchers to vary only selected features of the artifact while keeping the rest of the architecture constant. Ideally, these variations are controlled by parameters in the instantiated software itself. For example, researchers interested in the impact of information representation (e.g., tables vs. graphs) may, instead of producing different software systems, produce variations of tables and graphs by parameterizing this construct through software settings that control aspects such as the presence of headers or the presence of row highlights in the artifact.

Additionally, such parameterization addresses the recommendation of traceability from construct to features, in that it can be shown clearly how different levels of a construct affect the relevant features of the artifact [2]. And while such an instantiation cannot demonstrate that each feature represents only one construct, it allows the researcher to keep the remaining design (e.g., data or data model of the information system) constant, thus reducing the confounding influences on the feature.

Reusing existing instantiations is an ideal way to reduce artifact cost. However, a theoretical challenge is that artifacts in IS research are typically constructed for specific research questions, thus limiting their reusability in different circumstances. A pragmatic challenge of reusability is that artifacts created for research are not typically shared in the IS design research community. This issue comprises both legal aspects of licensing, as well as technical aspects of hosting source code or entire projects. To address the former, various types of licenses exist that differ in their degree of permissiveness, such as the GPL, BSD or Apache style licenses. To address the latter, open-source repositories, such as GitHub or SourceForge, exist to host a project. We encourage researchers to avail themselves of these options.

Artifact Instantiation Space. Most IS theories are moderately abstract (mid-range theories [24]). A challenge, therefore, is to account for the consequences of the chosen implementation and ensure that they do not interact with the variables of interest in unpredictable ways. Ideally, this requires a level of theoretical understanding of software construction that we currently do not possess. However, researchers should be able to identify at least some of the theories that relate to features of their artifact, and to explicate possible factors that could influence the artifact's features [25].

Once the relevant constructs are identified, researchers need to trace their effects on the instantiated features of the artifact to identify if, how, and under what conditions they affect the features. For example, a particular user interface button may be placed to enhance personalization, but at the same time the button may also increase complexity of interaction. This can confound any conclusions drawn from the study if it is unclear whether complexity has a causal effect on a dependent variable.

Another way to address the instantiation space is to construct multiple instantiations. If it is difficult to choose one valid instantiation or if there is conflicting theoretical guidance regarding a given property, a researcher can develop multiple artifacts, each corresponding to a different way of instantiating a construct. These different instantiations should behave identically with respect to the study's dependent variable (e.g., see [26]). This may be viewed as similar to convergent validity (all survey items should behave similarly) or predictive validity (valid survey items behave as expected with respect to a criterion variable) in survey research. Additionally, this can show the robustness of a theory to different implementations [24].

Artifact Complexity. In contrast to many simple experimental stimuli (e.g., line drawings [22]) or questionnaire items, a software system is a complex entity with many interacting parts. We recommend that user studies such as focus groups [17] and pretests be conducted during multiple stages of artifact design and development [9, 18] to identify confounding emergent properties early in the research cycle. To the extent possible, researchers need to ensure ceteris paribus equivalence across conditions when instantiating different levels of the focal construct [16]. When artifact features interact, especially through technical constraints, it may be that equivalence [27] cannot be established, in which case this threat cannot be addressed with experimental designs. It is important that researchers identify possible interaction effects and address them either through artifact design or the larger research design.

To reduce artifact complexity, we recommend that researchers should not embody the artifact with more features than required. While a minimal artifact may affect the ecological validity of the research, there is little to be gained by conducting work that lacks IV in a realistic setting. This also implies that researchers who employ existing artifacts should choose the "simplest" one, or limit users' exposure to the complete artifact, for example through training, disabling of functions, or access control.

Artifact Medium and Distance. In survey research the theoretical construct and the items intended to measure it are expressed in the same medium – natural language, so that validity may be established in part by using terminology from a construct's theoretical definition for questionnaire items. In contrast, instantiated IT artifacts are expressed in a different medium than the theoretical construct.

Part of this distance can be bridged by the creation of design principles, as proposed in [25, 28]. Such principles can make the focal theoretical construct more concrete. However, considerable distance remains to be bridged from design principles to instantiations [24]. We recommend making explicit the translation between the two media by establishing traceability in their research and demonstrating how the distance is bridged. Additionally, focus groups [17] at various stages can be used to ensure that the artifact design remains congruous with the theoretical definition of the construct (or at least the perception of the construct in the target population). Finally, parameterization of the software artifact can help impose a metric on the design space and thus allow a comparison with the theoretical space. For example, when personalization of the artifact can be controlled by varying a software parameter between the values 0 and 1, the resulting features show low and high personalization. These can then be compared with theoretical ideals of different personalization levels.

Technological Progress. IT artifacts continuously change their form and behavior due to relentless progress in computing power, emergence of new development methods, and new ways of interacting with systems. This means the validity of an instantiation may change over time and requires that validity be re-established or re-demonstrated for every instantiation that is used in a different way or context. Although, as noted earlier, we agree in principle with the idea of instrument re-use and standardization (in this case of software artifacts) [20], this can be challenging in the context of IS. For example, giving research participants a character and command-based interface to support their virtual collaboration in the age of sophisticated 3-D interfaces may create negative reactions not present when such interfaces dominated the IT landscape. We caution researchers in reusing of IS artifacts to evaluate the extent to which the context has changed and may no longer be appropriate.

4 Conclusion

The technology focus inherent in the IS discipline gives rise to unique validity challenges. Ignoring concerns related to instantiation validity can have significant negative consequences for knowledge contributions in IS and application of IS design knowledge in practice. This makes establishing principles of IV both important and urgent. In this paper we begin addressing the lack of established principles by turning to the IS discipline itself for guidance. A review of the IS discipline revealed valuable recommendations that we use to address the threats to IV. These recommendations can be used by researchers and reviewers in: testing; design science theorizing; evaluation of IT artifacts; and building real-world software systems based on IS design knowledge.

To ensure increased attention to the problem of IV, we recommend a new **"Instantiation Validity" section** be included in research papers. This section should link features of the artifact to the underlying theoretical constructs and attempt to demonstrate instantiation validity. Authors should take full advantage of different presentation modes made available by the publisher. With the growing popularity of online supplementary materials, researchers may share the artifact itself to aid other researchers, reviewers and practitioners assess the validity of the artifact.

This paper is an early attempt to generate guidelines for IV. We do not claim to provide a comprehensive set of recommendations. For example, we focused primarily on positivist aspects of IV, ignoring the relationship between artifact and the context in which it is built as well as how it may be interpreted by people in that context. Our intent was to initiate a dialogue and propose a direction that more comprehensive research on IV can take. Thus, future studies should look beyond IS and consider guidance from other domains with a design focus. Finally, as science in general, and standards of validity in particular, are socially agreed-on ideas, future work should engage IS researchers in a dialogue (e.g., using conference panels, survey and Delphi study methods) to expand, refine and prioritize guidelines for addressing threats to IV.

References

1. Komiak, S.Y.X., Benbasat, I.: The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q. 30, 941–960 (2006).
2. Lukyanenko, R., Evermann, J., Parsons, J.: Instantiation Validity in IS Design Research. Advancing the Impact of Design Science: Moving from Theory to Practice. pp. 321–328. Springer (2014).
3. Boudreau, M.-C., Gefen, D., Straub, D.W.: Validation in Information Systems Research: A State-of-the-Art Assessment. MIS Q. 25, 1–16 (2001).
4. Straub, D., Boudreau, M.-C., Gefen, D.: Validation guidelines for IS positivist research. Commun. Assoc. Inf. Syst. 13, 1–63 (2004).
5. Venable, J., Pries-Heje, J., Baskerville, R.: A comprehensive framework for evaluation in design science research. DESRIST. pp. 423–438. (2012).
6. Rossi, M., Sein, M.K.: Design research workshop: a proactive research approach. Present. Deliv. IRIS. 26, 9–12 (2003).
7. Hevner, A., Chatterjee, S.: Design Research in Information Systems: Theory and Practice. Springer (2010).
8. Sonnenberg, C., vom Brocke, J.: Evaluations in the science of the artificial–reconsidering the build-evaluate pattern in design science research. DESRIST. pp. 381–397. (2012).

9. Eriksson, C., Åkesson, M., Kautz, K.-H.: Authentic and Concurrent Evaluation: Refining an Evaluation Approach in Design Science Research. PACIS 2011 Proc. 1–13 (2011).
10. Purao, S.: Truth or dare: The ontology question in design science research. J. Database Manag. JDM. 24, 51–66 (2013).
11. Hevner, A., March, S., Park, J., Ram, S.: Design science in information systems research. MIS Q. 28, 75–105 (2004).
12. Walls, J.G., Widmeyer, G.R., Sawy, O.A.E.: Building an information system design theory for vigilant EIS. Inf. Syst. Res. 3, 36–59 (1992).
13. Gregor, S., Hovorka, D.S.: Causality: The elephant in the room in information systems epistemology. 19th Eur. Conf. Inf. Syst. (2011).
14. Chandra, L., Seidel, S., Gregor, S.: Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions. Hawaii Int. Conf. Syst. Sci. 4039–4047 (2015).
15. Venkatesh, V., Brown, S.A., Bala, H.: Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. MIS Q. 37, 21–54 (2013).
16. Benbasat, I.: Laboratory experiments in information systems studies with a focus on individuals: a critical appraisal. Harvard Business School, Cambridge, MA (1989).
17. Tremblay, M.C., Hevner, A.R., Berndt, D.J.: Focus Groups for Artifact Refinement and Evaluation in Design Research. Commun. Assoc. Inf. Syst. 26, 599–618 (2010).
18. Sein, M., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action design research. MIS Q. 35, 37 (2011).
19. Abraham, R., Aier, S., Winter, R.: Fail Early, Fail Often: Towards Coherent Feedback Loops in Design Science Research Evaluation. Int. Conf. Inf. Syst. 1–12 (2014).
20. Moore, G.C., Benbasat, I.: Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. Inf. Syst. Res. 2, 192–222 (1991).
21. Jarvenpaa, S.L., Dickson, G.W., DeSanctis, G.: Methodological issues in experimental IS research: experiences and recommendations. MIS Q. 141–156 (1985).
22. Snodgrass, J.G., Vanderwart, M.: A standardized set of 260 pictures: Norms for name agreement. J. Exp. Psychol. 6, 174–215 (1980).
23. Lukyanenko, R., Evermann, J.: A survey of cognitive theories to support data integration. AMCIS 2011 Proceedings. pp. 1–15. (2011).
24. Lukyanenko, R., Parsons, J.: Reconciling theories with design choices in design science research. Service-Oriented Perspectives in Design Science Research. pp. 165–180. (2013).
25. Arazy, O., Kumar, N., Shapira, B.: A theory-driven design framework for social recommender systems. J. Assoc. Inf. Syst. 11, 455–490 (2010).
26. Lukyanenko, R., Parsons, J., Wiersma, Y.: The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-generated Content. Inf. Syst. Res. 25, 669 – 689 (2014).
27. Burton-Jones, A., Wand, Y., Weber, R.: Guidelines for Empirical Evaluations of Conceptual Modeling Grammars. J. Assoc. Inf. Syst. 10, 495–532 (2009).
28. Kuechler, W., Vaishnavi, V.: A Framework for Theory Development in Design Science Research: Multiple Perspectives. J. Assoc. Inf. Syst. 13, 395–423 (2012).