Replication Crisis or Replication Reassurance: Results of the IS Replication Project

By: Alan R. Dennis  
Co-Editor, AIS Transactions on Replication Research  
Indiana University  
ardennis@indiana.edu

Susan A. Brown  
Co-Editor, AIS Transactions on Replication Research  
University of Arizona  
suebrown@arizona.edu

Taylor M. Wells  
Managing Editor, AIS Transactions on Replication Research  
Brigham Young University  
taylor.wells@byu.edu

Arun Rai  
Editor-in-Chief, MIS Quarterly  
Georgia State University  
arunrai@gsu.edu

Motivation and Objective of the IS Replication Project

Having confidence that findings reported in a peer-reviewed study published in a reputable journal is generalizable to some current or future context is critical for scientific progress (Open Science Collaboration 2015). If empirical results are not reproducible, how can we have confidence in the ability of our theories to explain and predict behavior? Research has shown that about one third or more of research published in leading psychology and science journals is not reproducible (Camerer et al. 2016; Camerer et al. 2018; Open Science Collaboration 2015). This has led to arguments that there is a replication crisis in Psychology (for discussions of this, see Camerer et al. 2018; Pashler and Wagenmakers 2012; Stroebe and Strack al. 2014).

Do we have a replication crisis in the Information Systems (IS) discipline? In 2014, Alan Dennis and Joe Valacich launched AIS Transactions on Replication Research (TRR) with a goal of having a dedicated outlet for replication research, thereby promoting replication in the IS discipline (Dennis and Valacich 2014). In October 2018, MIS Quarterly (MISQ) and TRR came together to launch the IS Replication Project, with the goal of replicating 25 articles published in MISQ and other top IS journals to understand the extent to which IS research is reproducible. We posted a call for participation on AIS World and invited all interested researchers to submit proposals for a replication study. We received 31 proposals, which ultimately turned into 21 papers published or accepted for publication in TRR by July 1, 2020.

In this editorial, we present the profile of the replication studies, analysis, and results comparing the replications to the original studies, and recommendations for replication research.

Profile of Studies in the IS Replication Project

The 21 replications represent 59 different authors from 29 different universities in the United States, Germany, China, Brazil, and Austria. Fourteen of the studies replicated work from MISQ, three from Information Systems Research, and one each from Journal of MIS, Journal of the AIS, Communications of the ACM, and Computers & Security. The majority of the original studies were surveys (16), although there were also four experiments and one Web scraping paper replicated. The methods used in the replications were consistent with the original study, with only one of the studies employing a minor adjustment (i.e., from field survey to online survey). (See Table 1.)

1Some of this is due to researcher fraud but the vast majority of cases are not (Stroebe and Strack 2014).
<table>
<thead>
<tr>
<th>Authors</th>
<th>Method</th>
<th>Location</th>
<th>Participants</th>
<th>N = Rep.</th>
<th>Technology Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 De Leoz &amp; Petter 2020</td>
<td>Choi et al. 2010</td>
<td>M</td>
<td>Survey</td>
<td>Survey</td>
<td>South Korea</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Ebrahimi &amp; Martinez 2019</td>
<td>Choi et al. 2015</td>
<td>M</td>
<td>Survey</td>
<td>Survey</td>
<td>SE Asia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exp.</td>
<td></td>
<td>Exp.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Erskine et al. 2020</td>
<td>Rutner et al. 2008</td>
<td>C</td>
<td>Survey</td>
<td>Survey</td>
<td>Not Reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Fischer et al. 2020</td>
<td>Benamati &amp; Lederer 2001</td>
<td>C</td>
<td>Survey</td>
<td>Survey</td>
<td>US</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Giddens &amp; Riemenschneider 2020</td>
<td>Rutner et al. 2008</td>
<td>C</td>
<td>Survey</td>
<td>Survey</td>
<td>Not Reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Hermes et al. 2020</td>
<td>Adjerid et al. 2018</td>
<td>M</td>
<td>Exp.</td>
<td>Exp.</td>
<td>US</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Ma et al. 2020</td>
<td>Van Slyke et al. 2006</td>
<td>M</td>
<td>Survey</td>
<td>Survey</td>
<td>US</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Mockus et al. 2020</td>
<td>Lindberg et al. 2016</td>
<td>C</td>
<td>Web Scraping</td>
<td>Web Scraping</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Moquin 2020</td>
<td>Ho et al. 2017</td>
<td>C</td>
<td>Survey</td>
<td>Survey</td>
<td>Multiple</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 Samhan &amp; Joshi 2019</td>
<td>Kim &amp; Kankanhalli 2009</td>
<td>C</td>
<td>Survey</td>
<td>Survey</td>
<td>Not Reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 Samtani et al. 2019</td>
<td>Johnston &amp; Warkentin 2010</td>
<td>M</td>
<td>Survey</td>
<td>Survey</td>
<td>US</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exp.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. IS Replication Project Studies
Table 1. IS Replication Project Studies (Continued)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Type*</th>
<th>Method</th>
<th>Location</th>
<th>Participants</th>
<th>N = Rep.</th>
<th>Technology Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shah &amp; Soror 2019</td>
<td>C</td>
<td>Survey</td>
<td>Survey</td>
<td>US</td>
<td>Undergrad students</td>
<td>Undergrad students</td>
</tr>
<tr>
<td>Terlizzi et al. 2019</td>
<td>C</td>
<td>Survey</td>
<td>Survey</td>
<td>Hong Kong</td>
<td>Respondents to a banner ad on a Hong Kong website</td>
<td>MTurk (US only)</td>
</tr>
<tr>
<td>Tourninho &amp; de Oliveira 2019</td>
<td>M</td>
<td>Survey</td>
<td>Survey</td>
<td>US</td>
<td>Undergrad students</td>
<td>Undergrad students</td>
</tr>
<tr>
<td>Yang et al. 2020</td>
<td>M</td>
<td>Survey</td>
<td>Survey</td>
<td>US</td>
<td>Home users (MTurk US only)</td>
<td>Organizational users (MTurk US only)</td>
</tr>
<tr>
<td>Young et al. 2020</td>
<td>C</td>
<td>Survey</td>
<td>Survey</td>
<td>Finland</td>
<td>Working professionals</td>
<td>IT professionals (Qualtrics sample)</td>
</tr>
<tr>
<td>Zeng et al. 2020</td>
<td>C</td>
<td>Field Survey</td>
<td>Online Survey</td>
<td>US</td>
<td>Household respondents</td>
<td>Undergrad students</td>
</tr>
</tbody>
</table>

Note: *Replication Type: (C)onceptual; (E)xact; or (M)ethodological.
† Samtani et al. (2019) used MTurk participants from the US (78%), India (16%), the Philippines (2%), and others (4%)
Exp = Experiment
Org. = Original Study; Rep. = Replication Study

The following is the breakdown by type of replication:

- **Exact** (exact copies of the context and methods of the original article): 2 articles
- **Conceptual** (exactly the same research questions or hypotheses, but use different measures, treatments, and/or analyses): 11 articles
- **Methodological** (exactly the same methods as the original study; i.e., measures, treatments, statistics etc. but conducted in a different context): 8 articles

**Analysis**

We tracked the details for each paper, paying particular attention to the differences between the original study and the replication. In addition to method deviations, we examined the number of years that had elapsed since the original study was published, the participants and locations for both studies, as well as the technology across the studies. Finally, we examined the degree to which the results of the replication matched the results of the original study, both for significant and nonsignificant results.

**Results**

The details of each replication study and its findings can be found as articles in TRR. A summary of the project’s findings is presented in Table 2. The results lead to an interesting pattern. Specifically, the degree of difference between the original paper and the replication is not the source of any systematic difference in the results. Replications with large deviations were as likely to replicate as those with small deviations. In short, IS does not have a replication crisis.

Tables 1 and 2 show that very few replications (10%) were exact replications in which the replication study’s method, participants, location, and technology matched those of the original study. Most replications varied two or more of these four design elements. The methods of the
replications almost always matched the methods of the original study, and just over half used the same technology (about 40% examined whether the theory generalized to a different technology). Most replications used a different type of participant and studied the theory in a different location.

The central focus of the Replication Project was to understand if replications generally find the same results as the original studies. The short answer is yes. The long answer is more nuanced. Most of the original studies in Table 1 tested hypotheses and used statistics to conclude if the hypotheses were supported or not. Most replication studies tested the same hypotheses, so we can examine the extent to which the findings from the replication match the findings from the original study. In other words, to what extent does the replication find the same hypotheses to be significant as the original and the same nonsignificant ones?

One useful metric for this is the F1 score; there are, of course, many others. We calculated the F1 score for each study and found that the average F1 score in our set of studies was 0.709. We view this as good, especially considering that the typical replication differed from the original study on at least two important dimensions.
We performed additional analysis to see if there were factors that predicted a replication study’s F1 score. For example, we expected that replications that more closely match the original study would have higher F1 scores, while those that replicate the study in a different location with a different technology and different participants would have lower F1 scores, especially those where much time has elapsed between the original study and the replication. We were wrong. We could find no such pattern; replications that differed on many dimensions from the original study were as likely to find high F1 scores as those that closely matched the original study.

Therefore, we conclude that good IS theories—those published in MISQ and other top IS journals—are likely to transcend time, technology, user, and location. We in IS do not face a replication crisis as our colleagues in other disciplines do. Instead, we have replication reassurance.

There are some individual replications that seem to deviate from the original more than others. As we examine these individual replications, we note two key elements that contribute to differences in these particular studies:

- **Technology**: In two of the three studies that are different from the overall set (i.e., the number of matching findings is far lower than the other replications), the technologies are quite different. Specifically, in one case the original technology was e-commerce, while the replication was social networking. In the second case, the original technology was PCs and PDAs, while the replication examined virtual reality. It is possible that there are distinctions between these technologies that call for additional theorizing to capture the contextual differences in technology and generalize the original theories from one technology to another.

- **User understanding of technology**: In one study, the original study used participants from MTurk (U.S. participants) and Prolific Academic (English first language participants), while the replication used German students. The focus of the studies is on privacy, with the original work published pre-General Data Protection Regulation and the replication after. Thus, one element to consider is a national policy change or shift in public perceptions as a contributor to the generalizability of research findings over time and between countries (e.g., pre- and post-COVID-19).

**Recommendations**

We offer the following recommendations to advance the field of IS:

First, we suggest a good way to develop understanding regarding replication is through incorporating replication as part of training doctoral students. The importance of replication and the advantages of having a replication culture that understands new empirical findings should be replicated are well known (Dennis and Valacich 2014). Replication is about examining the generalizability of a theory, not about questioning the validity of the original study. We assume that all studies are valid when published, so the goal of replication is to understand how theories can generalize to other times and contexts and what changes are needed to better enable them to do so. Replication is also an excellent way for students to learn research methods. The original paper provides the research question, relevant literature, theory, and/or data and systems information, plus a set of valid methods to answer the question. This allows students to focus on the method so that they can not only learn what it is, but how it is done, as well as the more subtle art of navigating the peer review process. Therefore, we suggest that replication should be a standard part of research methods education for all doctoral students, and as appropriate, master’s students as well.

Second, we recommend professional organizations and journal editors collaborate on encouraging and sharing the results of replication in aggregate, as we are doing here. Journals with different editorial objectives can collaborate to promote replication. While TRR routinely publishes replications, highly ranked IS journals are unlikely to. This Replication Project provides a model for the collaboration between a top journal like MISQ with a replication-dedicated outlet like TRR: replications of articles in MISQ are published in TRR and the results of the project are disseminated through a MISQ editorial.

Another example of such a collaboration is the badging project between TRR, the AIS journals, and MISQ. When a paper in one of these journals is replicated, the original paper receives a badge in the AIS digital library. The badge signals two very important things about the original study:

- It was important and interesting enough for another group of authors to replicate the study, usually in a different context.
- It provided sufficient contextual, methodological, and technical details to be replicated, an important issue we discuss below.

As the original paper in the AIS digital library links to the replication paper, scholars in the field can get a sense of how various elements (e.g., technology, participants) might influence the results. The replication study provides one more piece of evidence regarding the validity of the theory in the original study, which can be a useful reference when framing the arguments in a new piece of research.
Third, we suggest replication as an effective solution to assess whether the results are unique to one data set. To counter dubious methodological practices (e.g., data mining, HARKing, p-hacking) that may constrain results to one data set, some researchers advocate study preregistration, where researchers formally preregister their *a priori* theoretical hypotheses with a third-party and pledge not to deviate from this theory in their analyses after the data has been collected (Kupferschmidt 2018). Preregistration imposes theory-based constraints to solve a methodological problem. In contrast, replication directly addresses the question as to whether the results are generalizable beyond one data set.

Fourth, as sources of uncertainty in IS research can stem from novel data-generation processes, methods, and contexts (Rai 2020), authors need to include sufficient contextual, methodological, and technical detail to enable their work to be replicated, as we briefly discuss:

- **Contextual**: It is important that technology is sufficiently described along with other aspects of the context such as the temporal, location, cultural, and problem and task settings. This information will enable replication research to assess whether a change in context challenges assumptions underlying a theory or model; alters the meaning of constructs; changes the nature of causal, generative, or contingency mechanisms; or changes the utility of theories, models, or solutions.

- **Methodological**: Researchers should provide sufficient information on the protocols for data generation, analysis, and interpretation of results which involve:
  - Research design (e.g., sampling procedures, treatments and controls, procedures to source and curate data)
  - Analysis (e.g., validity assessment, model specifications, robustness tests)
  - Interpretation of results (including how researchers may have reconciled quantitative metrics on model accuracy and bias with human judgement)

  The specifics of the methodological information will depend on the particular methods that are employed. Where mixed-methods are employed, it is important to provide sufficient information on the individual methods and how they are combined in the study.

- **Technical**: Information on choices made in writing, modifying and executing code, and the scalability and portability of code and models, may be relevant to provide.

  From a pragmatic perspective, it is not possible, nor do we suggest, that authors try to provide every detail of every aspect of their projects in an article. Journal articles have page limits.

As authors may be constrained in providing sufficient detail to enable their work to be replicated, *TRR* has a special type of article that can be linked to the original article published in another journal to provide all researchers with access to the necessary empirical materials and protocols. This option can be a viable one for authors of all original research to consider in publishing their protocols, data, and/or code as a supplemental article in *TRR* which links to the original article in the other journal. The viability of this choice for an author will depend on the policies of the journal in which the authors have published their original article.

Finally, some researchers conducting replications want to replicate the original study exactly, while others prefer to use new methods to test the original study’s theory. Both are appropriate, although we note that replications using new methods are more powerful (Dennis and Valacich 2014).

**Concluding Remarks**

The results of the IS Replication Project are encouraging and provide replication assurance. Given our experiences with the project, we suggest that collaborative initiatives between replication-dedicated outlets and top scholarly journals, such as this initiative between *TRR* and *MISQ*, can play an important role in promoting replication and progress of the IS discipline. While we had an overwhelming majority of survey-based research in this project, we expect shared protocols, data, and code will enable greater replication across types of research and help with advancing the IS discipline.
References


