



The Trouble with Troubled Projects: Keeping Mum during Times of Crisis

Stacie Petter

Department of Information Systems
Baylor University
stacie_petter@baylor.edu

Julie DeJong

MiMedx Group, Inc.
julie.dejong@outlook.com

Adriane Randolph

Department of Information Systems
Kennesaw State University
arandolph@kennesaw.edu

Anthony T. Robinson

Department of Management and Marketing
Augusta University
atrobenson@augusta.edu

Abstract:

This study is a near exact replication of an experiment by Smith et al. (2001) that explores how individuals respond to the need to share bad news about an IT project. The replication results reinforce the crux of the original findings, which identify that when individuals think that negative information should be reported, they feel a greater responsibility to do so, but only to internal sources. In contrast with the original study, this replication study finds that a greater responsibility to report leads to more (rather than less) reluctance to report to an individual within the chain of command. Further, this study finds that an increased risk perception of a project does not necessarily result in a person being more likely to claim that the project's status ought to be reported, and perceived wrongdoing does not have a significant impact on responsibility to report. The results imply that whistle-blowing theories may vary greatly with internal versus external sources for reporting, and individuals may be sensitized by several notable company scandals and subsequent legislation to curtail wrongdoing. Further, researchers are encouraged to continue examining other factors impacting one's need and responsibility to report bad news.

Keywords: Mum effect, Escalation, IT project management, Reporting bad news, Project risk, Whistle-blowing

The manuscript was received 02/17/2016 and was with the authors 2 months for 1 revision.

1 Introduction

IT projects by nature include "high complexity, risk, and uncertainty" which creates many challenges in the process of IT project management (Keil and Montealegre, 2000). One challenge in the context of IT project management is avoiding runaway projects in which projects are over budget, over schedule, or may lack critical functionality. One reason for runaway projects is due to escalating commitment (Keil and Mann 1997), which occurs when individuals continue to pursue a project despite negative information that suggests continuing in that course of action may not be the best decision (Brockner, 1992; Staw and Ross, 1987).

In their original study, Smith et al. (2001) examined various factors that influence how an individual might report bad news about an IT project. In their original paper, Smith et al. (2001) offer two explanations why a person might remain mum (i.e., the "mum effect") and choose not to report bad news about an IT project. First, because of the complexities associated with IT projects, observers may lack confidence in their understanding of the severity of the problem and its impact. Second, the dynamic nature of projects causes directions and boundaries to change as time elapses, as new information is acquired, and as feedback affects the process. For this reason, observers may not be certain that their assessment of the project being in trouble is correct and subsequently may not decide to report the problem. The original study confirmed

these ideas along with recognizing that individuals are more likely to report negative information about an IT project to internal sources rather than external sources.

In this replication study, we study the same phenomenon of interest by examining how individuals respond to the need to share bad news about an IT project. In their original paper, Smith et al. (2001) noted the higher explanatory power of assessed personal responsibility when reporting news internally versus externally in that “if this finding were to be replicated in other studies, it would suggest that the central model, based largely on the whistle-blowing work of Dozier and Miceli [17], may be fully applicable only to *internal* reporting” (p. 209, emphasis in original). In the replication, the intention is to see if the results continue to hold and to identify whether there needs to be different theories of whistle-blowing to internal versus external sources.

Figure 1 shows the research model, paths, and results from the original study that are part of this replication. The hypotheses replicated from the original study are summarized in Table 1. The original study had multiple measures for reluctance to report to capture four different options for reporting to internal sources and three different options for reporting to external sources. The original study also measured perceived risk of negative personal consequences for reporting news to each source. Our replication measured the constructs shown below, as well as measured the same internal and external options for reporting the bad news. We did not include perceived risk of negative personal consequences for reporting news. Thus, H3a and H3b were not included in this replication study and are not shown in Figure 1 nor listed in Table 1 making this a nearly exact replication. Consistent with the original paper, the testing of H3c was accomplished outside of the PLS structural model using a non-parametric procedure.

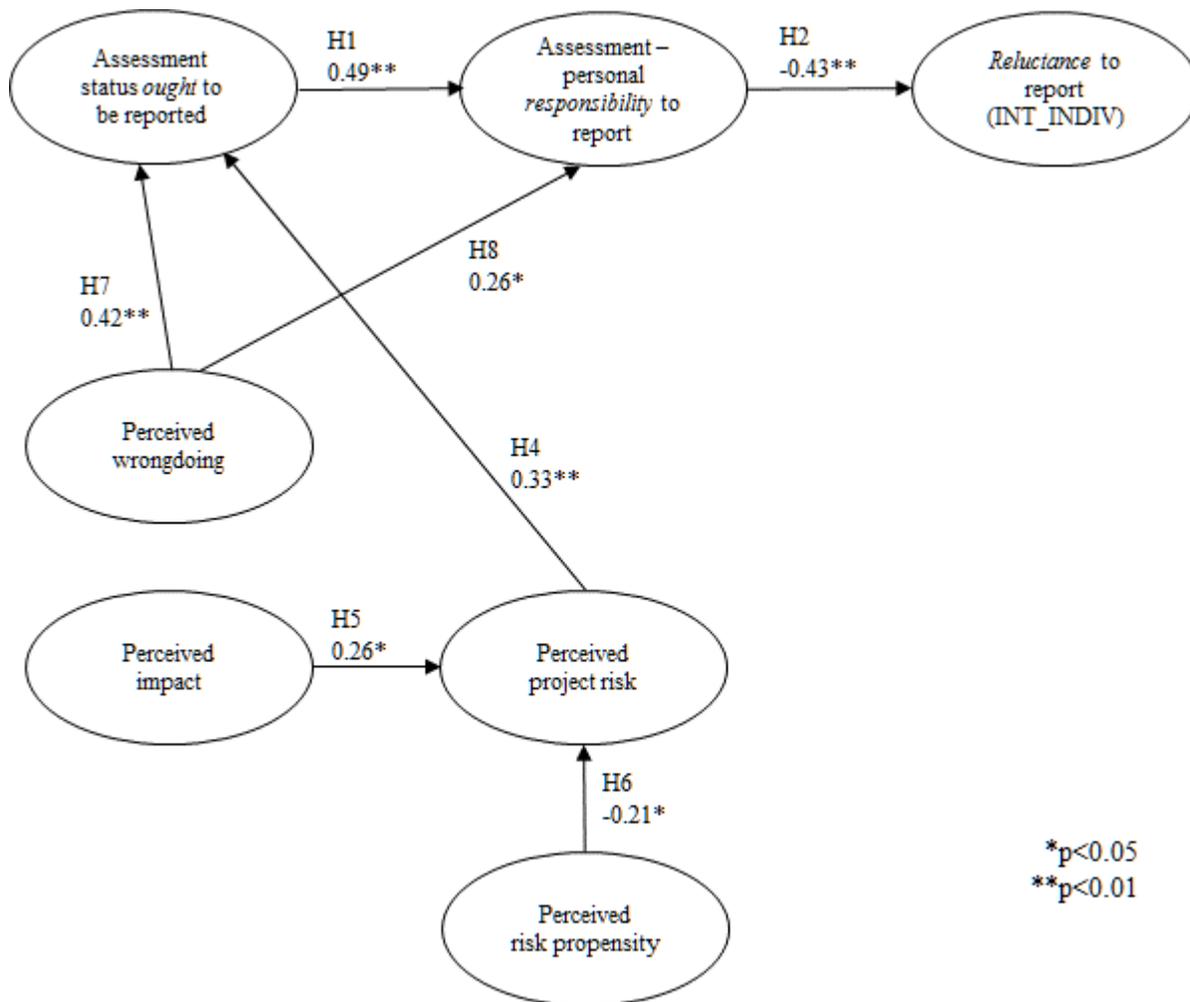


Figure 1. Research Hypotheses for the Replication Study and Results of the Original Study

Table 1. Hypotheses Replicated

H#	Hypothesis
H1	A greater perception that information ought to be communicated will be reflected in a higher level of perceived personal responsibility for reporting.
H2	The level of perceived personal responsibility for reporting will be inversely related to an individual's reluctance to transmit negative information. That is, higher levels of personal responsibility will be associated with less reluctance to report bad news.
H3c	Individuals will exhibit more reluctance to report through external than through internal reporting channels.
H4	Individuals will be more likely to claim that a project's status ought to be reported when perceived project risk is high.
H5	Perceived project risk will increase with perceived impact.
H6	There will be a significant negative relationship between perceived risk propensity and perceived project risk. For example, decision-makers perceiving themselves to have a risk-seeking propensity will perceive risks to be lower than decision-makers perceiving themselves to have a risk-averse propensity.
H7	Individuals are more likely to assess that negative information ought to be reported when they perceive higher levels of wrongdoing.
H8	Individuals are more likely to assess a personal responsibility to report a project's status when they perceive higher levels of wrongdoing.

2 Research Method

This study is a near-exact replication of Smith et al. (2001) in which we used a scenario-based experiment with a two-factor, four-cell design, consistent with the original study. We used the same scenarios and the same manipulations for perceived impact and level of wrongdoing, which were manipulated independently (again consistent with the original study). The scenarios asked subjects to read a short description of a software project that was in trouble, known as CAPS (see Appendix 1). Smith et al. (2001) developed four scenarios in which the level of impact had a high and low condition and the level of wrongdoing had a high and low condition. High impact scenarios stated that the CAPS project would have a major negative impact on the organization, while low impact scenarios suggested the project was not critical to the firm. High wrongdoing scenarios stated that the supervisor was planning to lie about the bad news, while the low wrongdoing condition explained that the supervisor was going to engage in honest behavior.

Consistent with the original study, subjects were randomly assigned to one of four treatment conditions. Each treatment condition consisted of a different scenario that manipulated impact and wrongdoing. After reading the scenario, subjects answered questions about the scenario, including manipulation checks. The subjects also had an opportunity to identify if they would respond to the situation with a range of alternatives, answered questions related to the research model, and provided demographic information. The same measurement items used in Smith et al. (2001) were used in this study.

Consistent with the Smith et al. (2001) study, this replication study used student subjects that were business students in a large urban university in the southeastern United States. Many subjects in our study at this university were non-traditional, so we asked about years of work experience. Subjects included business students taking a required introduction to information systems course as well as students in other required management courses. However, since the scenario contained terminology and a context that was specific to the information systems domain, we only retained participants for analysis that were either currently taking or had previously taken the required introductory information systems course. This ensured our population was consistent with the original study.

A total of 219 subjects participated in the study. Eight of the 219 surveys had significant levels of missing data, and these responses were not included in our data analysis. We also removed 51 subjects that had not taken the required introductory information systems course. This left 160 responses for analysis. Table 2 provides demographics for the remaining subjects in our study and compares the subject pool to those that participated in Smith et al. (2001).

Original Study		Replication Study	
Average Age	23.4 years	Average Age	24.0 years
Average Years of Work Experience	3.2 years	Average Years of Work Experience	6.5 years
% Female	55%	% Female	62%
N	163	N	160

Replication Notes:
The subjects in each sample were similar ages; however, the subjects in the replicated study had, on average, twice the work experience as the subjects in the original study. Our replication was comprised of a higher percentage of female students than the original study.

3 Results

3.1 Manipulation Checks

Two constructs were manipulated in this experiment through the text in the scenario: perceived impact and perceived wrongdoing. To evaluate the effect of the manipulation, each subject was asked four questions about the perceived impact of the project and the perceived wrongdoing of the supervisor. Consistent with the approach used by Smith et al. (2001), the measurement items for each construct were averaged. The means for perceived impact (1 = low level of impact; 7 = high level of impact) and perceived wrongdoing (1 = low level of wrongdoing; 7 = high level of wrongdoing) are illustrated in Figure 1 for each manipulation in the experiment. As shown in Table 3, the replicated results follow the same pattern found by Smith et al. (2001).

		Panel A				Panel B			
		Original Study				Replication Study			
Level of Wrongdoing (WRG)	High	N	40	41	N	39	42		
		Mean		Mean			Mean		
	Low	WRG	5.85	5.46	WRG	5.62	5.37		
		IMP	3.88	5.46	IMP	3.76	5.86		
		N	40	42			N	40	39
		Mean		Mean			Mean		Mean
		WRG	2.37	2.35			WRG	2.43	2.46
		IMP	3.57	5.95			IMP	3.16	5.80
		Low		High			Low		High
		Level of Impact (IMP)				Level of Impact (IMP)			

Replication Notes:
The number of subjects retained in our replication (after removing those that had not taken an introductory information systems course) is similar to the original study. Comparing Panel A with Panel B, the pattern of high/low wrongdoing/impact is also consistent with the original study, demonstrating that this aspect of the study was successfully replicated.

However, upon further analysis of the manipulation checks, we found that a large number of subjects did not successfully identify the manipulations within the scenarios. This deeper level of examination of the effectiveness of the manipulations was not described as having been performed in the original study. To evaluate if an individual passed a manipulation check, if a subject had a condition that was high (for impact or wrongdoing), the average value for the items measuring perceived impact or perceived wrongdoing

should have an average of 4.0 or higher since items were measured on a 7-point scale. If the subject received a condition that was low on perceived impact or perceived wrongdoing, the expected value for the average score for the construct should be 4.0 or lower. We found that 23% of our respondents failed either the manipulation check for impact or wrongdoing.

For those that failed the manipulation check, 14% of respondents failed due to not identifying the level of impact appropriately. Eleven percent (11%) of respondents failed to correctly identify the level of wrongdoing. Eliminating those respondents that failed the manipulation check left 123 responses for analysis. Table 4 compares the means for each of the experimental scenarios across the complete data set and the filtered data set in which the subjects that failed the manipulation check were removed. Examining Table 4, it becomes clear that the subjects had a difficult time identifying the low impact condition.

Table 4. Manipulation Check (Replication Study – Complete Data Set vs. Passed Manipulation Check)									
Panel A				Panel B					
Complete Data Set				Passed Manipulation Check					
Level of Wrongdoing (WRG)	High	N	39	42	Level of Wrongdoing (WRG)	High	N	24	34
			Mean	Mean				Mean	Mean
		WRG	5.62	5.37			WRG	5.96	5.81
		IMP	3.76	5.86			IMP	3.10	6.07
	Low	N	40	39		Low	N	29	36
			Mean	Mean				Mean	Mean
WRG		2.43	2.46	WRG	2.24		2.31		
	IMP	3.16	5.80		IMP	2.64	5.84		
		Low	High			Low	High		
		Level of Impact (IMP)				Level of Impact (IMP)			

Replication Notes:
 When using only the responses from subjects that successfully passed the manipulation check (Table 4, Panel B), the manipulations are stronger than the original study (i.e., Table 3, Panel A).

To further assess the effect of the manipulations, we performed the same 2x2 multiple analysis of variance (MANOVA) test as Smith et al. (2001). The average score for perceived impact and perceived wrongdoing were the dependent variables, and the manipulations were specified as the independent variables. In the original study, the main effects of each manipulated variable were strongly significant on the corresponding dependent variable and non-significant for the other dependent variable. In the original experiment, the impact manipulation was significant for perceived impact, but no significance was found between the impact manipulation and the wrongdoing manipulation. The wrongdoing manipulation was not significant in relation to perceived impact, but was found to be significant for perceived wrongdoing. The original study also found a “modest interaction effect” on perceived impact. Due to the large ratio difference between the sum of squares between the main effect and the interaction effect, this issue was considered to be minor in the original experiment.

In the replicated study, a 2x2 multiple analysis of variance was performed using the same approach, and similar results were expected. As Table 5 indicates, each independent variable was significant with the appropriate dependent variable in that the impact manipulation had a main effect on the impact composite variable and the wrongdoing manipulation had a main effect on the wrongdoing composite variable. Unlike the original study, there were no interaction effects identified in our replication. This suggests that the variables were successfully manipulated independently, which is an improvement over the original study.

Table 5. Results of 2x2 MANOVA (Original Study and Replication Study)				
Original Study Dependent Variable: Perceived IMPACT			Replication Study Dependent Variable: Perceived IMPACT	
Independent Variable	Sum of Squares	F-Value (Sig.)	Sum of Squares	F-Value (Sig.)
Main effect: IMPACT manipulation	160.664	117.600 (0.000)	222.150	35.613 (0.000)
Main effect: WRONGDOING manipulation	0.340	0.249 (0.619)	15.307	2.454 (0.120)
Interaction effect: (IMPACT manipulation) X (WRONGDOING manipulation)	6.440	4.714 (0.031)	1.846	0.296 (0.587)
Original Study Dependent Variable: Perceived WRONGDOING			Replication Study Dependent Variable: Perceived WRONGDOING	
Independent Variable	Sum of Squares	F-Value (Sig.)	Sum of Squares	F-Value (Sig.)
Main effect: IMPACT manipulation	1.764	1.265 (0.262)	21.173	1.001 (0.319)
Main effect: WRONGDOING manipulation	442.689	317.483 (0.000)	603.687	28.553 (0.000)
Interaction effect: (IMPACT manipulation) X (WRONGDOING manipulation)	1.388	0.995 (0.320)	29.380	1.390 (0.241)
<i>Replication Notes:</i> The original study found a slight interaction effect among the Impact and Wrongdoing manipulation when Perceived Impact was the dependent variable. The replication did not find any interaction effects for Perceived Impact or Perceived Wrongdoing. This implies that the replication was more successful than the original study at manipulating the two variables independently.				

3.2 Measurement Model Assessment

Before testing our hypotheses, we assessed reliability, convergent validity, and discriminant validity of the constructs. Reliability of the items in each construct was established using Cronbach's alpha. Ideally, each construct should have a Cronbach's alpha of 0.7 or higher to ensure "extensive" reliability (Bearden et al., 1993), yet three of the measurement constructs did not meet this minimum threshold. Nunnally (1967) states that constructs could have a Cronbach's alpha threshold of 0.6 if the work is exploratory. Yet, Perceived Project Risk did not meet this lower threshold for reliability.

We used factor analysis to assess convergent and discriminant validity. We first examined the outer loadings for each of the items to evaluate how well the items loaded onto each construct. Several items did not load well on their respective constructs (i.e., a loading < 0.6), and these items were dropped from the analysis. Specifically, one item from Ought to Report, one item from Perceived Risk Propensity, and one item from Perceived Project Risk did not properly load on their respective constructs. Table 6 contains a

summary of the reliabilities for each construct as well as a comparison of the item to construct loadings for the original and replicated study.

Table 6. Item-Construct Loadings (Original Study and Replication Study)					
Construct	Cronbach's Alpha		Item	Item to Construct Loading	
	Original	Replication		Original	Replication
Perceived Wrongdoing	0.90	0.93	WRG1 (reversed)	0.91	0.88
			WRG2	0.78	0.87
			WRG3 (reversed)	0.93	0.96
			WRG4 (reversed)	0.90	0.91
Perceived Impact	0.78	0.89	IMP1	0.79	0.91
			IMP2 (reversed)	0.77	0.80
			IMP3 (reversed)	0.81	0.90
			IMP4	0.75	0.86
Perception Status Ought to be Reported	0.67	0.63	O1	0.85	0.86
			O2 (reversed)**	0.76	0.55
			O3	0.74	0.76
Perceived Responsibility to Report	0.82	0.83	RSP1	0.87	0.93
			RSP2 (reversed)	0.83	0.81
			RSP3	0.87	0.84
Perceived Risk Propensity	0.77	0.60	RPRO1	0.75	0.71
			RPRO2**	0.66	0.44
			RPRO3	0.70	0.70
			RPRO4**	0.69	0.61 ¹
			RPRO5	0.76	0.67
Perceived Project Risk	0.74	0.58	RPRJ1	0.85	0.78
			RPRJ2 (reversed)	0.70	0.64
			RPRJ3	0.80	0.75
			RPRJ4 (reversed)**	0.58	0.57

*item dropped due to poor loading on construct

Replication Notes:

Perceived Wrongdoing and Perceived Impact had higher reliabilities and items had a higher loading on their respective constructs for the replication as compared to the original study. This is expected given that only subjects that passed the Perceived Wrongdoing and Perceived Impact manipulation check were retained in this study.

The construct, Perceived Responsibility to Report, consistently replicated from the original study to the new study.

The reverse coded items did not replicate well. In the replicated study, the reverse coded items for constructs not used in the manipulation had consistently lower loadings than the original study (which also affected the reliability in the replication).

Perceived Risk Propensity did not have any reverse coded items, yet this construct did not replicate well in this study.

To evaluate discriminant validity, we examined if the square root of the Average Variance Explained (AVEs) were greater than the correlations among the latent constructs. Consistent with the original study, our replication also passed this test for discriminant validity. Table 7 provides the squared correlations of the

¹ This item was also dropped from further analysis because the loading for RPRO4 fell below 0.6 once RPRO2 was removed from the construct. After removing RPRO2 and RPRO4 from the construct, the remaining item-to-construct loadings for the construct remained stable and greater than 0.60.

original study (upper half of matrix) and replication study (lower half of matrix) along with a comparison of the average variance extracted across studies.

	Average Variance Extracted		Squared Correlations among Latent Constructs <i>(original study in upper right half of matrix; replication study in lower left half of matrix)</i>					
	Original	Replication	WRG	IMP	O	RSP	RPRO	RPRJ
Wrongdoing	0.78	0.82	--	0.00	0.20	0.23	0.00	0.01
Impact	0.61	0.75	0.01	--	0.01	0.01	0.01	0.07
Ought	0.60	0.73	0.07	0.02	--	0.37	0.03	0.13
Responsibility	0.73	0.74	0.06	0.02	0.30	--	0.01	0.01
Project Risk	0.55	0.54	0.01	0.00	0.00	0.00	--	0.05
Risk Propensity	0.51	0.53	0.14	0.10	0.03	0.06	0.07	--

Replication Notes:
 In the replication, the values for Average Variance Extracted are equivalent to or higher than the values in the original study.

However, the squared correlations among latent constructs are generally lower in the replication as compared to the original study. The strongest correlation in both studies is the relationship between the perception that something ought to be reported and perceived responsibility to report. The authors of the original study only reported the squared correlations, which is why squared correlations are provided here for comparison purposes (NOTE: In the replicated study, any squared correlation that is 0.03 or higher is significant).

3.3 Structural Model

We evaluated the structural model using Smart PLS 3.0 (Ringle et al., 2015). Separate models were created to evaluate how the model might change as the dependent variable (Reluctance to Report) was altered for each reporting source. Figure 2 compares the results of the path coefficients and significance from the original study (O) with the results of the replication (R). In the original experiment, all hypotheses were supported with the exception of H2 for the three external options for reluctance to report. In the replication, we had similar results for H2 except that one of the internal options for reluctance to report was significant, but in opposite direction as expected. In the replication, H4 and H8 were not supported.

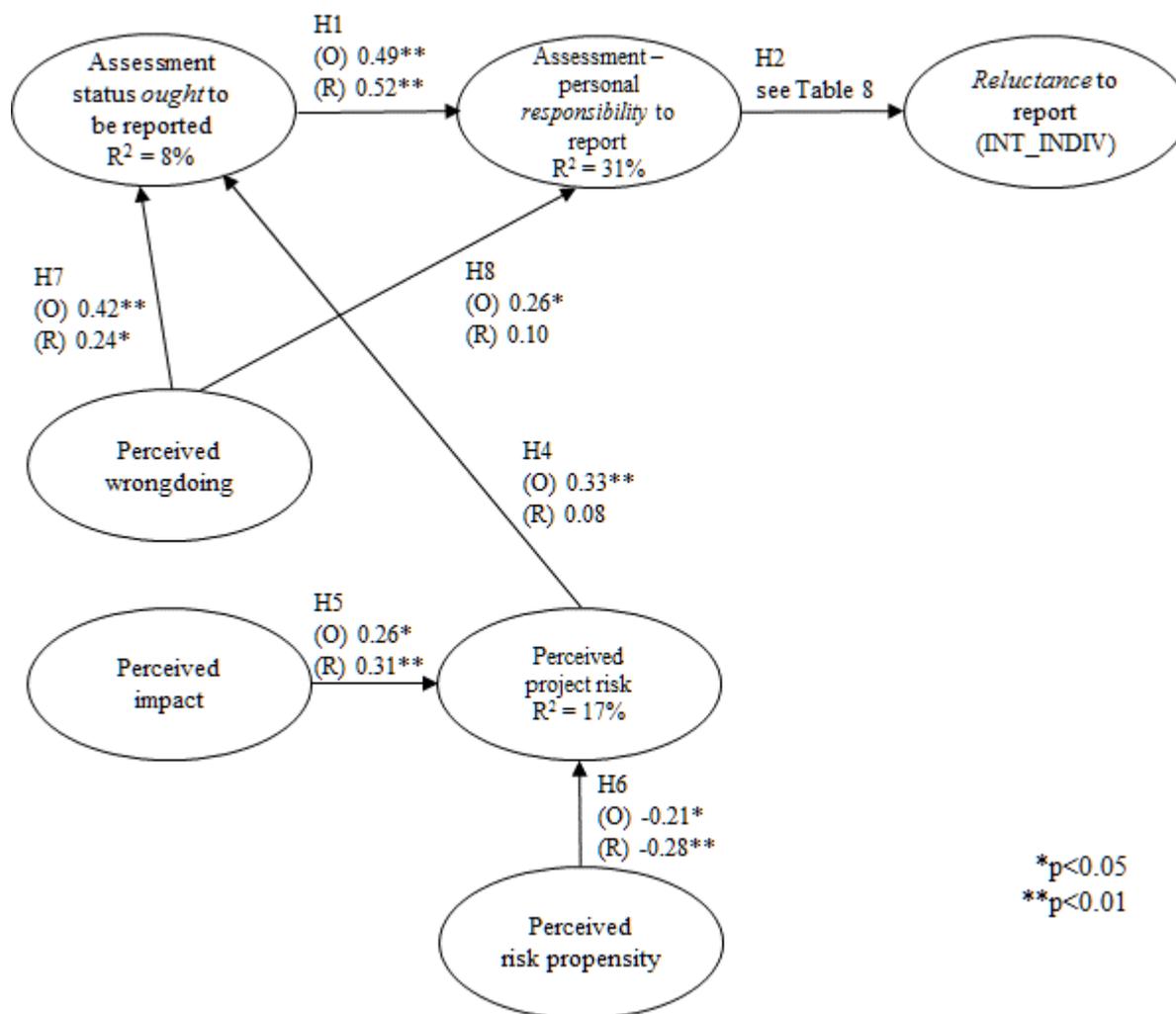


Figure 2. Results of the Original Study and Replication Study

Table 8 identifies the results of each of the hypothesis tests for various reporting alternatives. Each row corresponds to a different option for reporting the bad news about the IT project. Our experiment yielded both similar and different results as compared with the original study in regards to the relationship between responsibility to report and reluctance to report (H2). Consistent with the original study, H2 was not supported for any of the external reporting options. The path coefficient for H2 was significant for all internal reporting options. The reporting options of reporting to the group, expressing concerns to an auditor, and discussing the issue using the CIO’s open door policy had the expected significant negative relationship between responsibility and reluctance to report. Yet, the relationship between responsibility and reluctance to report to an individual had a *positive*, significant relationship. In our replication, the sign of the path coefficient for H2 for the individual reporting option was in the opposite direction as the results obtained by Smith et al. (2001).

Reporting Alternative	Original			Replication		
	Coefficient for H2	T-Value	R ²	Coefficient for H2	t-Value	R ²
Internal – Individual	-0.43	-4.41**	24%	0.25	2.87**	6%
Internal – Group	-0.27	-2.42**	24%	-0.23	2.69**	5%
Internal – Auditor	-0.21	-1.82**	23%	-0.32	3.72**	10%
Internal – Open Door	-0.22	-1.96*	15%	-0.28	3.00**	8%
External – Association	-0.18	-1.45	20%	-0.05	0.80	0%
External – TV	-0.04	-0.39	14%	0.02	0.43	0%
External – Client	-0.13	-1.29	18%	0.07	1.02	1%

**p < 0.01; *p < 0.05

Table 9 summarizes the replicated hypotheses and whether they were found to be supported or not. The core model of Smith et al. which considered the relationships among the perception that bad news *ought* to be reported, the perception the individual has a *responsibility* to report, and *reluctance* to report had similar outcomes as the original model (i.e., H1 supported and H2 partially supported). However, the variance explained in our replication was much lower for responsibility and reluctance to report.

In the original model, perceived wrongdoing significantly influenced ought to report (H7) and responsibility to report (H8). In the replication, perceived wrongdoing did predict ought to report (H7 supported), but did not influence responsibility to report (H8 not supported).

The original model suggested that perceived impact and perceived risk propensity influenced perceived project risk (H5 and H6). Then, perceived project risk influenced ought to report (H4). Our replication found support for the same relationships among perceived impact, perceived risk propensity, and perceived project risk (H5 and H6 supported). Further, the variance explained for perceived project risk was similar in both the original and replication. However, perceived project risk was not a significant predictor in ought to report (H4 not supported).²

To test H3c, we used a procedure similar to Smith et al. (2001). This hypothesis proposed that individuals would be more reluctant to report through external channels than through internal channels. Individuals that identified themselves as “likely” or “highly unlikely” to report using the alternative were identified as “reluctant.” We examined if each individual had a higher reluctance to report bad news to external sources. Then, we identified the proportion of subjects that had a higher reluctance for reporting bad news to external sources (proportion = 0.58). Using a one-tailed hypothesis test to compare the proportion of subjects with higher reluctance to report to external sources to the null hypothesis of 0.5 (equal chance of reporting internally or externally), we confirmed that H3c is supported ($p < 0.05$), consistent with the original study.

² Since Smith et al. (2001) did not report if they dropped any subjects from their analysis if they failed the manipulation check, we performed the same analyses using all of the data in which subjects that failed the manipulation check were retained in the data set. The primary results were identical to the results reported above. The only difference is that in some circumstances H7 was not supported when using all subjects rather than just those that successfully passed the manipulation check.

Table 9. Hypotheses Replicated

H#	Hypothesis	Original Result	Replication Result
H1	A greater perception that information ought to be communicated will be reflected in a higher level of perceived personal responsibility for reporting.	Supported	Supported
H2	The level of perceived personal responsibility for reporting will be inversely related to an individual's reluctance to transmit negative information. That is, higher levels of personal responsibility will be associated with less reluctance to report bad news.	Supported	Partial Support
H3c	Individuals will exhibit more reluctance to report through external than through internal reporting channels.	Supported	Supported
H4	Individuals will be more likely to claim that a project's status ought to be reported when perceived project risk is high.	Supported	Not Supported
H5	Perceived project risk will increase with perceived impact.	Supported	Supported
H6	There will be a significant negative relationship between perceived risk propensity and perceived project risk. For example, decision-makers perceiving themselves to have a risk-seeking propensity will perceive risks to be lower than decision-makers perceiving themselves to have a risk-averse propensity.	Supported	Supported
H7	Individuals are more likely to assess that negative information ought to be reported when they perceive higher levels of wrongdoing.	Supported	Supported
H8	Individuals are more likely to assess a personal responsibility to report a project's status when they perceive higher levels of wrongdoing.	Supported	Not Supported

As a post-hoc analysis to further examine reasons why there might be differences in the original study and the replication, we examined if there was potential for other systematic differences within our results. Recognizing that our subjects attended a university that is comprised of both traditional and non-traditional students, we found that the timing of the class in which the survey was administered had an effect on the results. Some subjects were solicited from classes that were offered in the evening (i.e., 7:15pm), while other subjects were solicited from classes offered in the morning (i.e., 11:00am or earlier). We found that the average age and average work experience was higher for those students in the evening courses. This is consistent with universities that cater to a large non-traditional student population in which some students are taking courses in the evening while working full time. We also noted using independent sample t-tests that there were differences among responses in several of the items, not just demographic variables. Three of the reporting options were different among those sampled while taking an evening class versus those taking a morning class. Students taking an undergraduate course in the evening (i.e., older with more work experience) were less reluctant to report the situation to an auditor (internal reporting) or to a neighbor that is a television reporter (external reporting), but were more reluctant to report the situation to the client (external reporting) than students taking morning classes (i.e., younger with less work experience). Students that were taking evening classes also rated specific items for Ought to Report (i.e., O2) and Responsibility to Report (i.e., RSP1 and RSP2) higher than students taking morning classes.³

Given that class time seems to be related to Age and Work Experience, we examined if these two variable influenced Reluctance to Report. We found that Age was negatively related to reluctance to report to an Auditor, while Work Experience was positively, significantly related to reluctance to reporting to an Auditor. Age and Work Experience were not significantly related to any other reporting options.

³ We also examined the structural models separately for those taking classes in the morning versus those taking classes at night. There is little change in the structural model between the results when examining the full sample (i.e., 123 subjects) and only those taking morning classes. Some of the hypotheses are no longer supported if we run the structural model with only the students taking evening classes. However, there is not sufficient power within this subsample to know if class time is the true moderator or if it is something else. This is discussed further in the next section.

4 Discussion

This study confirms the main goal of the replication by examining the impact of internal versus external reporting options for an individual sharing bad news about an IT project. The central model of decision making based on Dozier and Miceli's (1985) model examining the relationship of assessed project status that *ought* to be reported to the felt *responsibility* to report to an individual's *reluctance* to report was not supported for the external options for reporting. Yet, the results of our replication study confirm the crux of the original study and its applicability to internal reporting for most of the internal reporting options. This perhaps explains why subsequent conceptual replication studies by the authors only focus on internal reporting practices as they examine the influence of agency, timing, and culture (Keil et al., 2004; Park et al., 2008, 2009; Tan et al., 2003).

In the original study, the authors found that the strongest path coefficient between responsibility and reluctance was for reporting to an individual. In the replication study, the strongest path was for reporting to an auditor. This difference could be due to the high-visibility accounting scandals that took place in the early 2000's by Enron, Tyco, Worldcom, and Freddie Mac. Closely related, in 2002, the Sarbanes-Oxley Act (regulations that influenced accounting practices) was passed and became mandatory for publicly traded companies. These events altered curriculum in many business schools in the United States. Thus, some of the differences in the original study and the replication study might be due to the increased attention on accountability that is part of the business school curriculum. These results present an opportunity for future research as to why people choose different reporting options even among internal sources and examine how current events might influence one's choice to remain mum or blow the whistle if an unethical practice is occurring within an information systems project.

While the original study found the expected negative relationship between responsibility and reluctance to report to an individual, our study identified a strong, significant positive relationship between responsibility and reluctance to report to an individual. Essentially, this finding suggests that when an individual's personal responsibility to report was high, they were not likely to report the news to an individual. In this scenario, the individual is Steve Rice, who is the VP of Development and the supervisor of the project manager that was committing the wrongdoing. We are uncertain as to why we obtained this unexpected result. One potential reason for this is that respondents may not have wanted to break the "chain of command" in such a formal way. Reporting the wrongdoing to an auditor or via the open door policy might have been considered a more reasonable or appropriate approach for reporting unethical behavior. This may also be a further reflection of living in "post-Enron" times where distrust of an organizational chain of command for reporting may be lingering.

Interestingly here, we did not find a relationship to connect project risk to the central decision making model, indicating that individuals will think a project's status ought to be reported if the perceived project risk is high, as was found in the original study. In addition, perceived wrongdoing did not have a significant impact on responsibility to report. Other studies have indicated that more may be at play for an individual assessing whether they ought to report or feel responsible for reporting bad news about an IT project particularly related to organizational climate (Keil et al., 2004; Keil et al., 2010; Park and Keil, 2009). Perhaps unassessed perceptions about information asymmetry (Keil et al., 2004) or anonymity (Keil et al., 2010) and how tightly the project was being monitored muted perceptions about project risk. It is also possible that respondents had trouble discerning if there was a climate of silence (Park and Keil, 2009) in place when presented with a scenario of high project risk and professed lying by the project manager. This conflicting result signals a need for qualitative assessment to better understand the reasoning behind responses.

For an unknown reason, some of the respondents did not take to the treatment to one or both of the experimental variables, wrongdoing and impact. We did not find any significant differences among gender, class, class time, work experience, or any other systematic reason why not all subjects responded appropriately to the manipulation. It is unclear why the subjects did not identify the low impact condition as low, but the failed manipulation could be caused by the use of student subjects with little motivation to perform the task well or caused by factors related to the scenario itself. Perhaps this result is somehow tied to the scenario itself because subsequent studies using a similar experimental tool also reported several respondents not responding to treatment at the rates of 17% (Park et al., 2008) and 19% (Park and Keil, 2009). The results of this replication are consistent with findings from these other conceptual replications. These numbers indicate that wording needs to be adjusted in the experimental scenario to better help subjects discern differences in the treatment levels.

Through our post hoc analysis, we noted that this research on the mum effect might be quite sensitive to the population under study. While the sample of the original study and replication study had similarities in age and both studies had more than 50% females in the sample, there were differences in work experience among the samples. We found that there were differences in results when comparing respondents that answered the survey during a night class and respondents that answered the survey during a morning class. While age and work experience differed across these two groups, there might be other differences not accounted for in the demographic information collected. Years of work experience, age, and gender have since been found to have significant impact on whistle-blowing intention, where years of work experience and female gender have been positively associated with whistle-blowing intention while age has been negatively associated with whistle-blowing intention (Keil et al., 2010). In our replication, we had similar findings with work experience and age as predictors of reluctance to report to an auditor. The original study and replication study used undergraduates as their subjects; therefore, a study using working professionals might find different results. This opens up new opportunities for research to examine how other demographic factors, personal experiences, or other individual factors might influence how individuals respond to making the decision to blow the whistle on a troubled project.

In addition to these findings, we also found more general lessons learned about conducting replication studies. First, we found the importance of conducting a pilot study even though the previous study was conducted, externally reviewed, and published in one of the leading outlets for the field. We realized there was an opportunity missed by not exactly replicating the original study and including perceived risk of negative personal consequences for reporting news. Because it was not directly related to the purpose of our replication, we made the decision to exclude this variable; further, we did not conduct a pilot test to ascertain the true impact of its omission ahead of running our full study nor did we fully comprehend its complete representation in the original study from the published instrument. Thus, we were unable to learn if it still held a direct effect on reluctance to report. It seems the value of still including this variable would have been better realized if we had conducted a full pilot study ahead of this replication work. Hence we highlight a risk of replication studies that may jump past this step to re-test the protocols and model by taking at face value the published instrument. Further, a pilot might have revealed the potential measurement issues with some of the items that had to be removed in the study. We could have added additional items to avoid having constructs with only two measurement items (e.g., ought to report).

Second, we found that replications offer opportunities for extension by updating the instrument to reflect current options and discourse. In our case, we realized another missed opportunity by not updating the examples for external reporting to include social media outlets which arguably are a combination of publicity-related information paths and inter-organizational information paths by incorporating elements of being both news reporting and customer-focused. Going forward, subjects may expect options for reporting via Twitter or Facebook instead of a developer association hotline, or to a blogger than to a TV reporter. Again, we recommend that such options be vetted via a pilot study. This could have allowed us to perform both an exact replication as well as a conceptual replication within the same data set (Dennis and Valacich 2015). In future replications of this research, researchers may offer additional experimental scenarios to test alongside the existing ones. Although this adds considerable effort and additional subjects to test the content and face validity of these new scenarios, their addition would help offer more options to researchers than the same set of inputs.

5 Conclusion

This study sought to confirm findings presented by Smith et al. (2001) on project escalation and the “mum effect.” Consistent with the original study, this study showed that when a person thought a bad project should be reported (ought to be) that they felt a greater responsibility to do so. Further, with this higher level of responsibility felt, a person was more apt to report bad news about the project internally but did not necessarily have a propensity to report to a source external to the company. People with risk-seeking propensities perceived risks to be lower than those with risk-averse propensities. Also, in light of wrongdoing, people thought they ought to report bad news about the project. Although perceived project risk increased with perceived impact as in the original study, this increased perception of risk did not consistently result in a person being more likely to claim that the project’s status ought to be reported, in contrast with the original study.

Thus, this nearly exact replication study largely supports the findings of the original study. It demonstrates that much of the central decision making model is consistent with a new sample. Yet, this replication opens up new areas for consideration and future research. Given some of the problems with the measurement items, this replication suggests that researchers interested in re-using these items and constructs in future studies should perform pilots and potentially alter the measures to ensure they perform well given the sample and context. Further, the results spotlight an opportunity for future research to more closely examine the effects of internal and external reporting sources that was described in the original study but not yet taken up in the field. It seems that even amongst the consistent findings in support of internal reporting behaviors, there are subtle differences regarding the actual role of the person in whom an individual confides. In addition, this replication study supports the ongoing work by the original authors to examine additional individual, organizational, and managerial factors impacting an individual's perception of responsibility to report and his/her ability to assess that there is even something worth reporting.

Acknowledgements

The authors would like to acknowledge the encouragement and support of Mark Keil, Detmar Straub, and Wynne Chin in conducting this replication study.

References

- Bearden, W.O., Netemeyer, R.G. & Mobley, M.F. (1993). *Handbook of marketing scales*, Newbury Park, CA: Sage.
- Brockner, J. (1992). The escalation of commitment to a failing course of action: Toward theoretical progress, *Academy of Management Review*, 17(1), 39-61.
- Dennis, A.R. and Valacich, J.S. (2015). A replication manifesto. *AIS Transactions on Replication Research*, 1, 1-4.
- Dozier, J. B., & Miceli, M. P. (1985). Potential predictors of whistle-blowing: A prosocial behavior perspective. *Academy of Management Review*, 10(4), 1985, 823-836.
- Keil, M. & Mann, J. (1997). The nature and extent of IT project escalation: Results from a survey of IS audit and control professionals (Part 1). *IS Audit & Control Journal*, 1, 1997, pp. 40-48.
- Keil, M. & Montealegre, R. (2000). Cutting your losses: Extricating your organization when a big project goes awry. *Sloan Management Review*, 41(3), 55-68.
- Keil, M., Mixon, R., Saarinen, T. & Tuunainen, V. (2004). Understanding runaway information technology projects: Results from an international research program based on escalation theory. *Journal of Management Information Systems*, 11(3), 65-85.
- Keil, M., Smith, H. J., Pawlowski, S., & Jin, L. (2004). 'Why didn't somebody tell me?': Climate, information asymmetry, and bad news about troubled projects. *Database for Advances in Information Systems*, 35(2), 65-84.
- Keil, M., Tiwana, A., Sainsbury, R., & Sneha, S. (2010). Toward a Theory of Whistleblowing Intentions: A Benefit-to-Cost Differential Perspective. *Decision Sciences*, 41(4), 787-812.
- Nunnally, J.C. (1967). *Psychometric Theory*, New York: McGraw Hill.
- Park, C., Im, G., & Keil, M. (2008). Overcoming the mum effect in IT project reporting: Impacts of fault responsibility and time urgency. *Journal of the Association for Information Systems*, 9(7), 409-431.
- Park, C., & Keil, M. (2009). Organizational silence and whistle-blowing on IT projects: An integrated model. *Decision Sciences*, 40(4), 901-918.
- Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. Bönningstedt: SmartPLS. Retrieved from <http://www.smartpls.com>.
- Smith, H.J., Keil, M., & Depledge, G. Keeping mum as the project goes under: Toward an explanatory model. *Journal of Management Information Systems*, 18(2), 189-227.
- Staw, B. M., & Ross, J. (1987). Knowing when to pull the plug. *Harvard Business Review*, 65(2), 68-74.
- Tan, B. C. Y., Smith, H. J., Keil, M., & Montealegre, R. (2003). Reporting bad news about software projects: Impact of organizational climate and information asymmetry in an individualistic and a collectivistic culture. *IEEE Transactions on Engineering Management*, 50(1), 64-77.

Appendix A: Experimental Scenario (from Smith et al. 2001)

Case Instructions: The business case that follows is part of a study that examines business decision-making. Please take a few minutes to read over the case and to answer the questionnaire that follows. There are no right or wrong answers.

Computer Diagnostics Corporation

You work for Computer Diagnostics Corporation (CDC), a computer software company. At any time, CDC has several state-of-the-art development projects in progress.

You are the lead systems analyst on one such project, CAPS—a project that has been underway for eight months and which is scheduled to be completed four months from now. Your direct supervisor is the project manager, Ron Gibbons. Ron is fully responsible for both task and personnel management on the project, and in turn, reports to the vice president of development for the company, Steve Rice. A partial organization chart below shows the reporting relationships and where you fit within the hierarchy.

President

|

Steve Rice Vice President, Development

|

Ron Gibbons Project Manager

|

YOU ARE HERE – Lead Systems Analyst

|

Development Team

Next week, CDC must make a financial investment in new hardware specifically for this project. Because of this, Steve Rice has asked for a status report. In preparation, you and Ron met for three hours yesterday. During that time, it became apparent that the project is in deep trouble on two critical dimensions:

1. Because the CAPS project has pushed the technological boundaries more than the typical project, there have been many unexpected technical glitches. Thus, when you and Ron took a careful look at the progress on all the tasks, you realized that the programmers had fallen badly behind schedule on several critical items.
2. A recent meeting with some potential customers convinced you that the user requirements had been partially misunderstood. Therefore, it is unclear whether the product that is currently being developed will be successful in the marketplace.

You and Ron performed a crude risk assessment of the project and concluded it had a 65% probability of failure given all the risk factors. You were both unpleasantly surprised and wished that you had examined the situation in greater depth at an earlier date.

You both know that the investment in CAPS so far represents 90 percent of the overall development budget for the current year, and the project is critically important to the company. Should this project fail, it would have a significant negative effect on the company's financial position, forcing it to layoff half its employees and to file for bankruptcy.

You both know that the project is potentially in jeopardy. ***You asked Ron how he planned to handle the situation with Steve Rice, and he replied "We can't let Steve know anything about this, even if it means lying about the status and the risk. He'll clean our clocks. Our position is that 'everything***

is coming along fine at this point.' Don't say anything negative to anybody, and shred all these papers we have scattered around the room.

You remain quite concerned about the project and are wondering what to do, if anything, about bringing your concerns to others, either within or outside of CDC.

The above scenario contains the treatment used in the *high impact* and *high wrongdoing* scenario. For the treatment of *low impact* and *high wrongdoing*, the following paragraph was substituted for the first bolded paragraph in the scenario above.

You both know that the investment in CAPS so far has been much less than 1% of the overall development budget for the current year. You also know that the project is not strategically important to the company. In fact, should this project fail, it would have no effect on the company's financial position.

For the *low wrongdoing* treatment, the following paragraph was substituted for the second bolded paragraph in the original scenario.

You both know that the project is potentially in jeopardy. ***You asked Ron how he planned to handle the situation with Steve Rice, and he replied "I've never lied to Steve before, and I won't start now. I'm going to call him immediately to report that we may not be on track. I'll give him a copy of the task status analysis that we did."***

.

Appendix B: Measurement Instrument

Reluctance to Report Bad News (7-point scale with 1 = very likely; 7 = very unlikely)

Variable	Original		Replication		Item
	Mean	s.d.	Mean	s.d.	
INT_INDIV	3.68	1.81	4.38	1.60	How likely are you to go directly to Steve Rice by yourself to discuss the status of the CAPS project?
INT_GROUP	3.77	2.08	3.28	1.59	How likely are you to try and persuade members of the development team to go to Steve Rice as a group to discuss the status of the CAPS Project?
INT_AUD	4.27	1.88	4.19	1.60	Suppose that CDC has an internal auditor who is responsible for ensuring that company policies and procedures are followed. This auditor reports directly to the CEO. How likely are you to tell the internal auditor about the status of the CAPS project?
INT_OPNDR	3.33	1.75	3.33	1.66	Suppose that the Computer Diagnostics Corporation has an "open door" policy under which an employee can speak with any senior executive—up to and including the CEO—whenever the employee wishes to do so. How likely are you to use the "open door" policy to discuss the status of the CAPS project with a senior executive?
EXT_ASSOC	5.47	1.53	4.98	1.69	Suppose that a professional association of software developers, of which you are a member, has instituted a "hotline" that its members can call to report incidents of "inappropriate management behavior" associated with software development projects. Since the association promises to investigate each report it receives, it is unlikely that your identity can be kept secret. How likely are you to call this hotline and discuss the status of the CAPS project?
EXT_TV	6.10	1.24	1.98	1.25	Suppose that a local television station has reported that CDC has "sharply rising expenses in the software development area." One of your neighbors is the lead financial reporter for the television station and has asked you to comment. Because your relationship with the reporter is well known, it would be almost impossible to hide your identity. How likely are you to discuss the status of the CAPS project with the television reporter?
EXT_CLNT	4.68	1.78	4.53	1.75	Suppose that one of CDC's major clients is placing a high level of reliance on the scheduled availability of CAPS because they plan to include it in a major new product announcement. An acquaintance of yours works for the client as a senior new product development manager on that project. How likely are you to discuss the status of the CAPS project with the client's new product development manager?

Perception of Whether Something Ought to Be Reported (7-point scale with 1 = strongly disagree; 7 = strongly agree)

Variable	Original		Replication		Item
	Mean	s.d.	Mean	s.d.	
O1	4.75	1.85	5.28	1.27	I believe that something should be done to make more information about the status of the CAPS project known to others.
O2 (reversed)	2.94	1.77	2.80	1.41	I don't believe that it really matters whether more information about the status of the CAPS project is made known to others.
O3	5.04	1.83	5.32	1.46	Even if it is not me, I believe someone should tell others about the status of the CAPS project.

Perceived Responsibility to Report (7-point scale with 1 = strongly disagree; 7 = strongly agree)

Variable	Original		Replication		Item
	Mean	s.d.	Mean	s.d.	
RSP1	4.69	1.68	5.04	1.38	I believe that – as the lead systems analyst on the CAPS project team – I have a personal responsibility to make more information about the status of the CAPS project known to others.
RSP2 (reversed)	3.36	2.00	2.99	1.51	I believe that it is not my responsibility – a lead systems analyst – to make more information about the status of the CAPS project known to others.
RSP3	4.13	1.85	5.01	1.51	I believe that it is my personal duty as CAPS's lead systems analyst to tell others about the project's status.

Perceived Impact (7-point scale with 1 = strongly disagree; 7 = strongly agree)

Variable	Original		Replication		Item
	Mean	s.d.	Mean	s.d.	
IMP1	4.70	1.97	4.61	2.10	I believe the degree of harm that could result from the CAPS project's failure is very high.
IMP2 (reversed)	2.50	1.76	2.77	1.71	In the end, I don't believe that it really matters whether the CAPS project fails or succeeds.
IMP3 (reversed)	3.13	2.02	3.17	1.89	I believe that, even if the CAPS project fails, there is little harm that could come from it.
IMP4	3.87	2.17	3.78	2.18	I believe that CDC's financial future rides on the CAPS project.

Perceived Wrongdoing (7-point scale with 1 = strongly disagree; 7 = strongly agree)

Variable	Original		Replication		Item
	Mean	s.d.	Mean	s.d.	
WRG1 (reversed)	3.74	2.30	3.84	2.15	I believe that Ron has managed the CAPS project in a very ethical manner.
WRG2	3.59	2.27	3.67	2.19	I believe that Ron's actions in yesterday's meeting were immoral.
WRG3 (reversed)	4.01	2.43	4.04	2.27	I believe that Ron has done the right thing in how he plans to handle Steve Rice.
WRG4 (reversed)	3.87	2.21	3.96	2.11	I believe that Ron's actions in yesterday's meeting are a good example of values-based leadership.

Perceived Risk Propensity (7-point scale with 1 = Extremely LESS likely than others; 7 = Extremely MORE likely than others)

Variable	Original		Replication		Item
	Mean	s.d.	Mean	s.d.	
RPRO1	3.37	1.30	3.70	1.36	Your tendency to choose risky alternatives based on the assessment of others on whom you must rely
RPRO2	3.66	1.36	3.95	1.43	Your tendency to choose risky alternatives that rely upon analyses high in technical complexity
RPRO3	3.76	1.54	3.95	1.45	Your tendency to choose risky alternatives which could have a major impact on the strategic direction of your organization
RPRO4	3.08	1.37	3.24	1.40	Your tendency to initiate a strategic corporate action which has the potential to backfire
RPRO5	2.83	1.62	2.58	1.22	Your tendency to support a decision when you were aware that relevant analyses were done while missing several pieces of information.

Perceived Project Risk (7-point scale anchors for each item)

Variable	Original		Replication		Item
	Mean	s.d.	Mean	s.d.	
RPRJ1	4.85	1.47	4.72	1.45	How would you characterize the current status of the CAPS project at Computer Diagnostics Corporation? (left anchor = significant opportunity; right anchor = significant threat)
RPRJ2 (reversed)	2.79	1.49	3.39	1.48	How would you characterize the current status of the CAPS project at Computer Diagnostics Corporation? (left anchor = potential for loss; right anchor = potential for gain)
RPRJ3	5.17	1.40	4.96	1.80	How would you characterize the current status of the CAPS project at Computer Diagnostics Corporation? (left anchor = positive situation; right anchor = negative situation)
RPRJ4 (reversed)	3.12	1.33	3.43	1.54	What is the likelihood that the CAPS project will be successful? (left anchor = very unlikely; right anchor = very likely)

About the Authors

Stacie Petter is an Associate Professor of Information Systems in the Hankamer School of Business at Baylor University. She received a Ph.D. in Computer Information Systems from Georgia State University and an M.B.A. with a concentration in Computer Information Systems from Georgia State University. She has published her research in journals such as *MIS Quarterly*, *Journal of Management Information Systems*, *Journal of the Association for Information Systems*, and *European Journal of Information Systems*, among others. She serves as an associate editor for *MIS Quarterly* and *Information Systems Journal*. Her primary research interests include IT project management, impacts of information systems, and research methods.

Adriane Randolph is the founder and executive director of the BrainLab and an Associate Professor of Information Systems in the Michael J. Coles College of Business at Kennesaw State University. She earned a Ph.D. in Computer Information Systems from Georgia State University and a B.S. in Systems Engineering with Distinction from the University of Virginia. Her research of over fourteen years focuses on brain-computer interface systems, which allow for non-muscularly controlled assistive technologies and reflect varying cognitive states. Other research interests include human-computer interaction and neuro-information systems. She has been featured as a speaker for multiple TEDx events and was invited to be an original Google Glass Explorer. She has published manuscripts in the *International Journal of Human-Computer Interaction* and the *ACM Transactions on Accessible Computing*. Prior to academia, she worked for Accenture, implementing change management and human performance tools in the federal government sector.

Julie DeJong earned her Master's Degree in Decision Sciences from Georgia State University's J. Mack Robinson College of Business. She has taught nearly ten years as a managerial sciences instructor at Georgia State University. She has over twenty years of experience in management decision support including "Big 5" consulting, insurance and healthcare billing and reimbursement compliance, and university managerial sciences training. She specializes in data analytics, business intelligence, project management, and process improvement.

Anthony T. Robinson holds a Ph.D. in Strategic Management, a Master of Business Administration, a Master of Science in Organizational Change, and a Bachelor of Business Administration in Finance. He serves on the faculty in the Hull College of Business at Augusta University and teaches in the areas of management, marketing, and entrepreneurship. His research tends to leverage a cognition lens in the areas of management, entrepreneurship, and innovation. He has published articles and presented at numerous international conferences (where his research has received Best Paper Awards). Some titles of his research include: "Risk Perceptions, Venture Creation Decisions, Overconfidence and Perceived Environmental Munificence"; "Entrepreneurial Expert Scripts: Does the Way Entrepreneurs Learn Matter for Venture Performance?"; "Founder Overconfidence: The Paradox of Overconfident Entrepreneurs"; "Technological Drivers of Innovation Activity in Entrepreneurial Firms"; and "The Interaction Between Market Cycles, Technological Change and Resource Configurations Among Alliances to Predict Performance."

Copyright © 2015 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from ais@aisnet.org.