Using WarpPLS in E-Collaboration Studies: What if I Have Only One Group and One Condition?

Ned Kock, Department of International Business and Technology Studies, Texas A&M International University, Laredo, TX, USA

ABSTRACT

What if a researcher obtains empirical data by asking questions to gauge the effect of an e-collaboration technology on task performance, but does not obtain data on the extent to which the e-collaboration technology is used? This characterizes what is referred to here as a scenario with one group and one condition, where the researcher is essentially left with only one column of data to be analyzed. When this happens, often researchers do not know how to analyze the data, or analyze the data making incorrect assumptions and using unsuitable techniques. Some of WarpPLS's features make it particularly useful in this type of scenario, such as its support for small samples and the use of data that does not meet parametric assumptions. The main goal of this paper is to help e-collaboration researchers use WarpPLS to analyze data in this type of scenario, where only one group and one condition are available. Two other scenarios are also discussed – a typical scenario, and a scenario with one group and two before-after technology introduction conditions. While the focus here is on e-collaboration, the recommendations apply to many other fields.

Keywords: *Action Research, Electronic Collaboration, Field Research, Multivariate Statistics, Partial Least Squares, Structural Equation Modeling, WarpPLS*

INTRODUCTION

Occasionally a researcher will obtain empirical data by asking questions that attempt to gauge the effect of an e-collaboration technology on task performance. However, the researcher will not obtain data on the extent to which the e-collaboration technology is used.

While this scenario may look like the result of bad research design, it is a relatively common scenario in investigations where the researcher gains access to the participants by offering a service to them, as is frequently the case in action research investigations (Kock, 2005; 2010b).

When this happens, frequently the researcher does not know how to analyze the data, or analyzes the data using unsuitable techniques. Some of WarpPLS's features make it particularly useful in this type of scenario, such as its support for small samples and the use of data that does not meet parametric assumptions.

DOI: 10.4018/jec.2013070101

The main goal of this paper is to help e-collaboration researchers use WarpPLS to analyze data in this type of scenario, where only one group and one condition are available. Two other scenarios are also discussed. These two scenarios are discussed to set the stage for the discussion of the one group and one condition scenario.

The first is a typical e-collaboration study scenario in which the researcher measures the degree to which the e-collaboration technology is used, or the degree to which specific features of the e-collaboration technology are used, as well as team performance and/or related variables expected to be influenced by e-collaboration technology use. This is a cross-sectional data collection scenario, and is one of the most common scenarios in e-collaboration research.

In the second scenario the researcher does not have data on the extent to which the e-collaboration technology is used, but has data related to team performance and/or other variables expected to be influenced by e-collaboration technology use before and after the technology is introduced. This is a longitudinal data collection scenario, and is a relatively common scenario in e-collaboration research.

The focus of this paper is on e-collaboration research, but the techniques discussed apply to a wide variety of fields. In fact, they arguably apply to any field in which attitudinal and/or behavioral responses are studied in connection with change, where change may result from the use of a new technology, from the introduction of a new management technique, from the use of a new drug for treatment of a disease, or even from a change in weather.

A TYPICAL E-COLLABORATION STUDY SCENARIO

Let us assume that a researcher introduced an e-collaboration technology into an organization with the goal of facilitating the work of business process improvement teams (Kock, 2005). These are teams that carry out business process redesign projects – they select, analyze and redesign business processes (Kock, 2006).

All teams studied by the researcher use the e-collaboration technology. No controls on how much the teams use the technology are applied by the researcher, characterizing the investigation as a field study with quasi-experimental elements (Shadish et al., 2002). The researcher is interested in the possible effect that the use of the e-collaboration technology has on team performance measures, such as the return on investment of a business process redesign project.

In this scenario, the researcher can measure the degree to which the e-collaboration technology is used, or the degree to which specific features of the e-collaboration technology are used. Either way, the researcher will have one or more variables for which there will be different values for different teams. These values will reflect different degrees of use of the e-collaboration technology as a whole, or of specific features of the technology.

The researcher can next collect team performance measures and build one or more models to be analyzed with WarpPLS (Kock, 2010; 2011; 2011b). A simple model would have two latent variables, one measuring e-collaboration technology use and the other measuring team performance, with e-collaboration technology use pointing at team performance.

Figure 1 shows an example dataset with data collected from 20 business process improvement teams. “ECU” measures the degree to which a team used an e-collaboration technology, on an 11-point scale from 0 to 10. “Perf” measures each team’s performance in the business process improvement task, also on an 11-point scale from 0 to 10.

Figure 2 shows a simple two-variable model, with results, implemented with WarpPLS using a linear analysis algorithm. In this model “ECU” is the predictor and “Perf” the

Figure 1. Example dataset for typical scenario

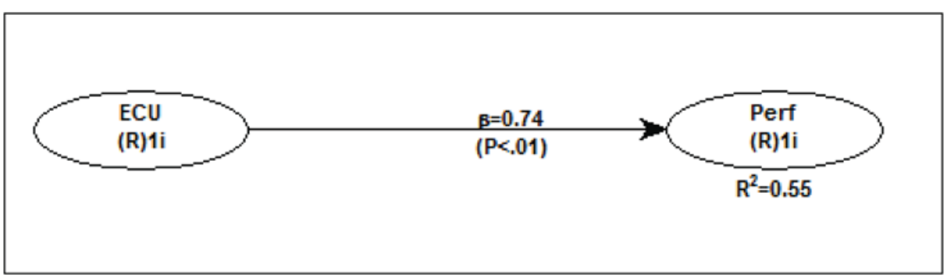
	A	B
1	ECU	Perf
2	3	0
3	7	6
4	1	4
5	4	2
6	10	10
7	0	0
8	10	10
9	4	2
10	6	9
11	8	4
12	2	4
13	3	2
14	4	4
15	8	5
16	8	6
17	3	3
18	1	2
19	7	10
20	10	6
21	4	4

Notes:
ECU = the degree to which the team used the e-collaboration technology
Perf = the team's performance in the business process improvement task

criterion. For simplicity, each latent variable is measured through a single indicator, and thus is not a "true" latent variable. Each indicator refers to the single corresponding column in the dataset.

Based on these results, we can conclude that the use of the e-collaboration technology was significantly and positively associated with the performance of business process improvement teams (β=0.74, P<.01). This significant associa-

Figure 2. Example WarpPLS model with results for typical scenario



Notes:
ECU = the degree to which the team used the e-collaboration technology
Perf = the team's performance in the business process improvement task

tion is a reflection of the fact that the variables "ECU" and "Perf" vary in concert with one another; i.e., these two variables are strongly and positively correlated. Since the model has only two variables, and a linear analysis is being conducted, the path coefficient for the association equals the correlation coefficient between the variables. That is: β=R=0.74, P<.01.

A good way of reporting the results of a typical scenario such as this, especially when only two variables are present, is to show a plot with the variables for "ECU" and "Perf", with an indication of the strength of the association and the level of statistical significance at the bottom of the plot. This is illustrated through Figure 3.

The plot shown has been directly generated by WarpPLS, copied, cropped with Paint (a picture editor made available by Microsoft with Windows), and pasted into this paper. WarpPLS also allows users to save plots as JPEG files (extension ".jpg") and then import them into papers.

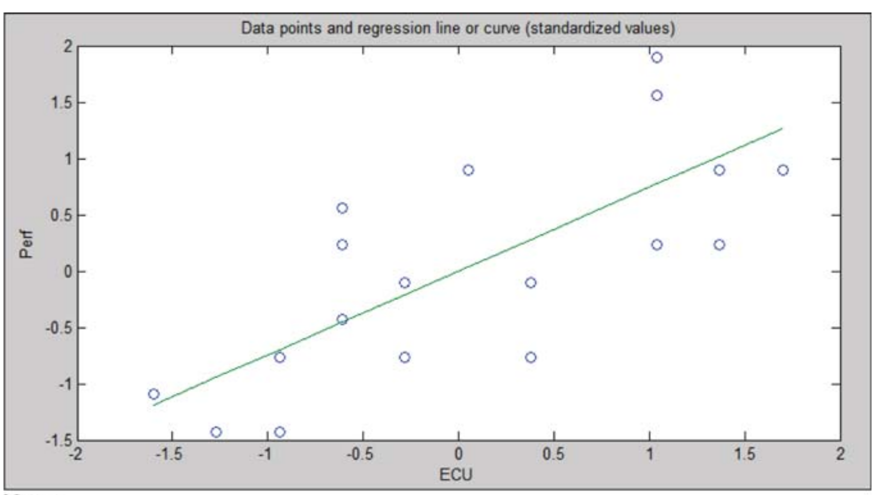
The underlying algorithm used was "PLS regression", a linear algorithm. Since only one-indicator latent variables were used, the "robust path analysis algorithm", a simpler algorithm, could also have been used to reach the same results. The values along the axes are standardized, e.g., 0 = mean, 1 = one standard deviation above the mean, -1 = one standard deviation below the mean, 1.5 = one and a half standard deviations above the mean, and so on.

A SCENARIO WITH ONE GROUP AND TWO CONDITIONS

What if the researcher does not have data on the extent to which the e-collaboration technology is used? In this case, the study has effectively only one condition. That is, a variable measuring e-collaboration technology use will have only one value; say "1", reflecting the state in which the technology "was used".

However, in this scenario, the researcher can collect data about team performance before and after the e-collaboration technology is introduced. This effectively creates two conditions to which the same group of people are subjected, which could be coded as "0" for "before the technology" and "1" for "after the technology".

Figure 3. Plot for WarpPLS model with results for typical scenario



Notes:
- ECU = the degree to which the team used the e-collaboration technology
- Perf = the team's performance in the business process improvement task
- ECU → Perf association statistically significant (β=0.74, P<.01)

Figure 4 shows an example dataset with data collected from 20 business process improvement teams for the scenario with one group and two conditions. "Bef0Aft1" codes for the before-after e-collaboration technology introduction condition, with "0" for "before the technology" and "1" for "after the technology". "Perf" measures each team's performance in the business process improvement task, also on an 11-point scale from 0 to 10.

Figure 5 shows a simple two-variable model, with results, implemented with WarpPLS. In this model "Bef0Aft1" is the predictor and "Perf" the criterion. As before, each latent variable is measured through a single indicator, for simplicity, and thus is not a "true" latent variable. Each indicator refers to the single corresponding column in the dataset.

This model conceptually implements the equivalent to a nonparametric comparison of means test, such as the Mann–Whitney U test. However, the latter would not allow for the use of latent variables, or introduction of covariates into the model, which are both made possible with WarpPLS. An equivalent parametric test would be an ANOVA test (Rosenthal & Rosnow, 2007).

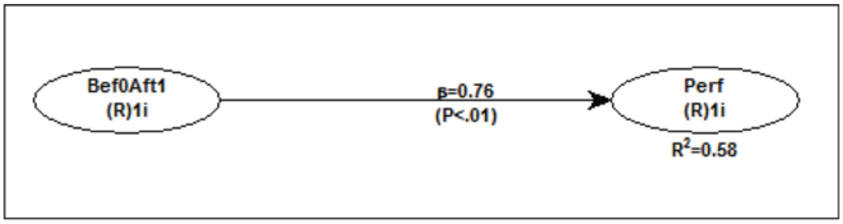
Based on these results, we can conclude that the use of the e-collaboration technology was associated with a significant increase in the

Figure 4. Example dataset for scenario with one group and two conditions

	A	B
1	Bef0Aft1	Perf
2	0	3
3	0	6
4	0	1
5	0	5
6	0	6
7	0	3
8	0	1
9	0	6
10	0	2
11	0	1
12	1	4
13	1	4
14	1	5
15	1	9
16	1	7
17	1	5
18	1	4
19	1	5
20	1	9
21	1	4

Notes:
- Bef0Aft1 = 0 (before the technology) and 1 (after the technology)
- Perf = the team's performance in the business process improvement task

Figure 5. Example WarpPLS model with results for scenario with one group and two conditions



Notes:
- Bef0Aft1 = 0 (before the technology) and 1 (after the technology)
- Perf = the team's performance in the business process improvement task

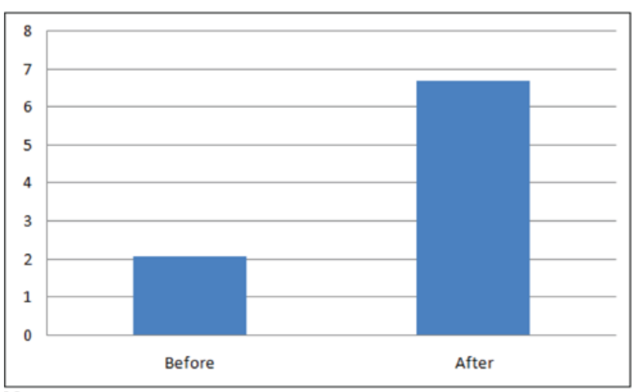
performance of business process improvement teams (β=0.76, P<.01). This increase refers to a comparison of performance measures before and after the introduction of the technology.

Probably one of the best ways of reporting the results of a scenario with one group and two conditions, before and after the introduction of the technology, is to show a bar chart with the mean scores before and after the technology, with an indication of the strength of the association and the level of statistical significance at the bottom of the chart. This is illustrated through Figure 6.

The mean scores for performance before and after the technology, reflected in the heights of the bars in the chart, can be easily calculated with many spreadsheet software tools, such as Microsoft Excel.

Is it not overkill to conduct a comparison of means analysis with WarpPLS? That is, instead of using a software tool that implements a simpler parametric technique such as ANOVA? Arguably not, because ANOVA and other related techniques require parametric assumptions to be met; e.g., that the criterion variable be normally distributed. This methodological argument has

Figure 6. Chart with means for scenario with one group and two conditions



Notes:
- Bar heights indicate the mean scores for performance before and after the technology
- Performance = the team's performance in the business process improvement task
- The difference between means is statistically significant (β=0.76, P<.01)

been demonstrated recently in a different context – a study of surprise-enhanced cognition (Kock & Chatelain-Jardón, 2011).

One of the advantages of using WarpPLS in a test like this is that it calculates P values using a nonparametric class of estimation techniques, namely resampling estimation techniques. Sometimes these techniques are referred to as bootstrapping techniques, which may lead to confusion since bootstrapping is also the name of a type of resampling technique. Unlike parametric techniques such as ANOVA, nonparametric estimation techniques do not require data to be normally distributed or large samples to yield credible results.

Another advantage of conducting a comparison of means analysis using WarpPLS is that the analysis can be significantly more elaborate. For example, the analysis may include control variables (or covariates), which would make the test equivalent to an ANCOVA test. Also, the analysis may include latent variables as predictors, criteria, or control variables. Technically speaking, this is not possible with ANOVA, ANCOVA, or commonly used nonparametric comparison of means tests such as the Mann-Whitney U test.

A SCENARIO WITH ONE GROUP AND ONE CONDITION

What if the researcher does not have data on the extent to which the e-collaboration technology is used, and, additionally, the researcher collected data about team performance only after the technology was introduced? In this case, we have only one group and one condition.

While this scenario may seem unlikely, it is in fact very common in investigations where the researcher gains access to the participants by offering a service to them, as often is the case in action research investigations (DeLuca et al., 2008). In these types of investigations the researcher often has little control on the subjects or the environment being studied (Kock, 2003), which may prevent data collection about team performance before and after the technology is introduced.

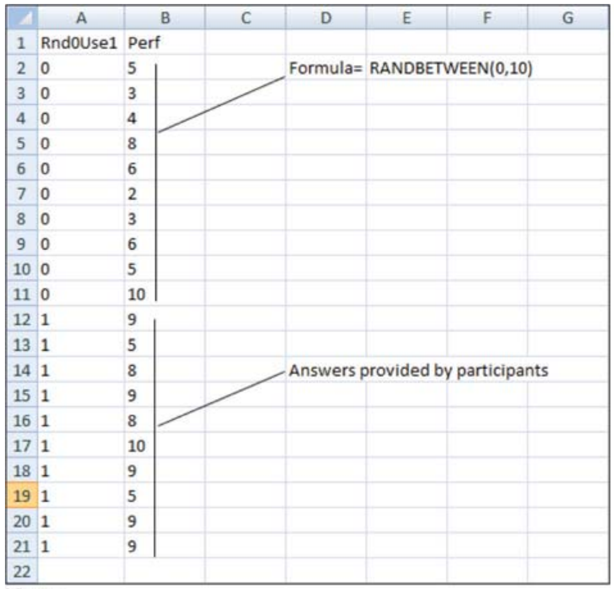
Within the context of e-collaboration use by business process improvement teams, this scenario would be typified by the researcher asking questions that attempt to gauge the effect of the technology on task performance, instead of asking questions that attempt to gauge task performance at a given point in time (as in the "one group and two conditions" scenario). One example would be the following generic question, answered on the 11-point scale shown in Figure 7: *In your opinion, what has the use of the e-collaboration technology done to the performance of business process improvement teams?*

Figure 8 shows an example dataset with data collected from 10 business process improvement teams for the scenario with one group and one condition. The dataset has 20 rows, for consistency with the previous examples. It has 20 rows because 10 of those rows contain random values created by the researcher. Although 10 rows of random values are being used here to be consistent with the previous example, the ability of WarpPLS to use data that does not conform to parametric assumptions means that we could generate 20, 30 or 100 rows.

Figure 7. 11 point answer scale

	0	1	2	3	4	5	6	7	8	9	10	
(Decreased it a lot)						(Had no effect)					(Increased it a lot)	

Figure 8. Example dataset for scenario with one group and one condition



	A	B	C	D	E	F	G
1	Rnd0Use1	Perf					
2	0	5		Formula=	RANDBETWEEN(0,10)		
3	0	3					
4	0	4					
5	0	8					
6	0	6					
7	0	2					
8	0	3					
9	0	6					
10	0	5					
11	0	10					
12	1	9					
13	1	5					
14	1	8		Answers provided by participants			
15	1	9					
16	1	8					
17	1	10					
18	1	9					
19	1	5					
20	1	9					
21	1	9					
22							

Notes:
- Rnd0Use1 = 0 (random value) and 1 (technology use answer)
- Perf = team performance increase in the business process improvement task

"Rnd0Use1" codes for whether the value is a random value or an answer in response to a technology use question such as the one above, with "0" for "random value" and "1" for "technology use". "Perf" measures each team's performance in the business process improvement task, also on an 11-point scale from 0 to 10.

The random values created by the researcher must be comparable with those obtained from answers in response to the technology use question. Since the latter were provided on an 11-point scale from 0 to 10, the random values must also be on an 11-point scale from 0 to 10. These can be created in Microsoft Excel with the formula "RANDBETWEEN(0,10)".

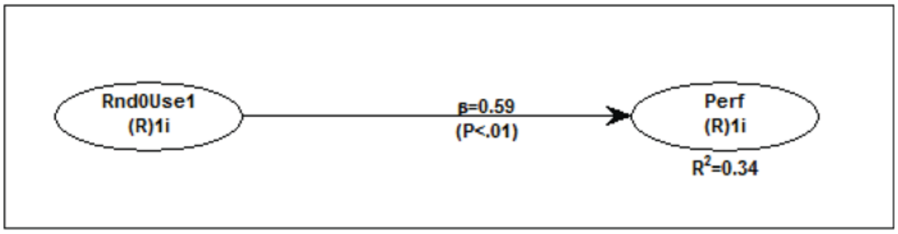
Figure 9 shows a simple two-variable model, with results, implemented with WarpPLS. In this model "Rnd0Use1" is the predictor and "Perf" the criterion. As before, each latent variable is measured through a single indicator, for simplicity, and thus is not a "true" latent variable. Each indicator refers to the single corresponding column in the dataset.

Based on these results, we can conclude that the use of technology was associated with a significant perceived increase in the performance of business process improvement teams (β=0.59, $P<.01$). This increase is evidenced by the answers given by the participants being significantly higher on average than random answers on an 11-point scale going from 0 to 10 (Figure 7). The random answers have a theoretical mean of 5, which on the 11-point scale provided would refer to the "no effect" point.

Probably one of the best ways of reporting the results of a scenario with one group and one condition is to show a bar chart with the mean scores for random (or chance) and technology use responses, with an indication of the strength of the association (i.e., the difference between means) and the level of statistical significance at the bottom of the chart. This is illustrated through Figure 10.

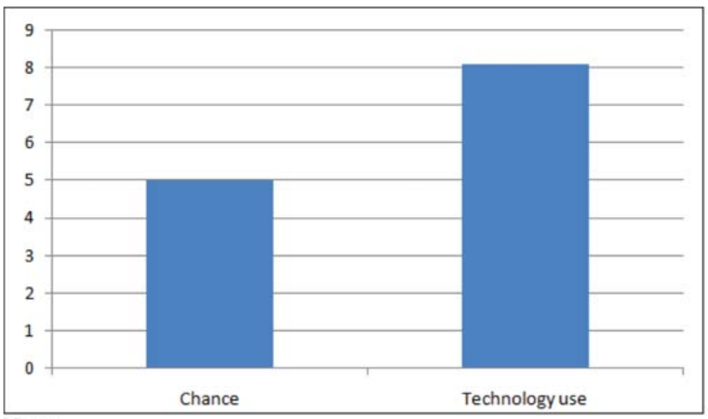
The sample of random answers had a mean score of approximately 5; the precise value was 5.2. Nevertheless, the height of the "chance" bar is indicated as 5, the theoretical mean. The

Figure 9. Example WarpPLS model with results for scenario with one group and one condition



Notes:
- Rnd0Use1 = 0 (random value) and 1 (technology use answer)
- Perf = team performance increase in the business process improvement task

Figure 10. Chart with means for scenario with one group and one condition



Notes:
- Bar heights indicate the mean scores for performance increase by chance and with the use of technology
- Performance increase = team performance increase in the business process improvement task
- The difference between means is statistically significant (β=0.59, P<.01)

reason why the sample mean was not exactly 5, the theoretical mean, is that we used a formula-generated sample of only 10 values, to match the 10 values regarding data from business process improvement teams. To avoid larger deviations from the theoretical mean, it may advisable to use larger samples of automatically generated random answers, and ensure that the mean is as close to the theoretical mean as possible.

For example, one may use 100 automatically generated random answers, recalculating them a few times until a mean that is very close to the theoretical mean is obtained. With Microsoft Excel, this can be done by pressing the "Shift" and "F9" keys simultaneously a few times; showing the mean of the automatically generated random answers in a cell with a formula such as "AVERAGE(random answers)" for easy reference.

Given the nonparametric resampling techniques employed by WarpPLS, there is no need to ensure that the "chance" and "technology use" samples have the same size. However, it is still important to ensure that the values in these two samples vary on the same scale. In this example, the scale is a Likert-type 11-point scale from 0 to 10.

DISCUSSION

One distinguishing feature of the scenario with one group and one condition, in comparison with the scenario with one group and two conditions, is that the data in the former is typically obtained through questions aimed at gauging the perceived effect of the e-collaboration technology on a criterion variable such as task performance. Below are a few examples:

In your opinion, what has the use of the e-collaboration technology done to the performance of business process improvement teams?

In your opinion, what has the use of the e-collaboration technology done to the cost of business process improvement teams?

In your opinion, what has the use of the e-collaboration technology done to the completion time of business process improvement teams?

The questions above refer to three criterion variables, or constructs, which could be called "team performance", "team cost", and "team completion time". The questions can be answered on a Likert-type scale, such as the one illustrated in Figure 7, or based on other scales, including ratio scales – e.g., team completion time reduction, measured in days.

How can the scenario with one group and one condition discussed earlier be extended to a situation with three criterion variables? In this case, the dataset would be similar to the one for the scenario discussed earlier, but would have four columns instead of two.

One of the columns would be "Rnd0Use1", coding for whether the value is a random value or an answer in response to a technology use question like one of the three questions above, with "0" for "random value" and "1" for "technology use".

The other three columns would refer to each of the criterion variables. For example, they could be labeled "Perf" for team performance, "Cost" for team cost, and "Time" for team completion time. All of the values under these three columns corresponding to the zeros under the "Rnd0Use1" column would be random-generated. The values under these three columns corresponding to the ones under the "Rnd0Use1" column would be based on the answers provided in response to the three questions.

The WarpPLS model would have "Rnd0Use1" as the single predictor in the model. The criteria variables in the WarpPLS model would be "Perf", "Cost" and "Time". That is, in this WarpPLS model, "Rnd0Use1" would point at "Perf", "Cost" and "Time".

Finally, the bar chart with the mean scores would have four bars. On the left would be a bar representing the mean chance score, which would be 5, if an 11-point scale going from 0 to 10 were to be used. Next to it would be bars with the mean scores for "Perf", "Cost" and "Time". These would allow one to visually contrast the mean scores for those three criteria variables with the use of technology with the mean chance score.

CONCLUSION

This paper's main goal is to help e-collaboration researchers using WarpPLS to analyze data in a scenario where only one group and one condition are available. This scenario would be typified by the researcher asking questions that attempt to gauge the effect of the technology on task performance, but without gathering data about the extent to which the technology is used. This is common in investigations where the researcher gains access to the participants by offering a service to them, as often is the case in action research investigations.

Two other scenarios are discussed to set the stage for the discussion of the scenario with one group and one condition. One is a typical e-collaboration study scenario, where the researcher measures the degree to which the e-collaboration technology is used, or the degree to which specific features of the e-collaboration technology are used, as well as team performance and/or other variables that are expected to be affected by e-collaboration technology use.

The other scenario includes one group and two conditions. In this scenario the researcher does not have data on the extent to which the e-collaboration technology is used, but collects data about team performance and/or other variables that are expected to be affected by e-collaboration technology use before and after the technology is introduced.

While the focus here is on e-collaboration research, the techniques discussed apply to a wide variety of fields. They apply to any field in which attitudinal and/or behavioral responses are observed after a change is effected, where the change may be the use of a new technology, the introduction of a new management technique, or even a change in an environmental condition surrounding individuals. The unit of analysis may be the individual or group, including large groups such as entire organizations or even nations.

ACKNOWLEDGMENT

The author would like to thank Aditya Limaye and James Cox for their comments and suggestions on an earlier version of this paper. The author is also grateful to the members of the PLS-SEM e-mail distribution list for questions, comments and discussions on topics related to the use of the software WarpPLS.

REFERENCES

DeLuca, D., Gallivan, M., & Kock, N. (2008). Furthering IS action research: A postpositivist synthesis of four dialectics. *Journal of the Association for Information Systems*, *9*(2), 48–72.

Kock, N. (2003). Action research: Lessons learned from a multi-iteration study of computer-mediated communication in groups. *IEEE Transactions on Professional Communication*, *46*(2), 105–128. doi:10.1109/TPC.2003.813164.

Kock, N. (2005). *Business process improvement through e-collaboration: Knowledge sharing through the use of virtual groups*. Hershey, PA: Idea Group Publishing. doi:10.4018/978-1-59140-357-9.

Kock, N. (2006). *Systems analysis and design fundamentals: A business process redesign approach*. Thousand Oaks, CA: Sage Publications.

Kock, N. (2010). Using WarpPLS in e-collaboration studies: An overview of five main analysis steps. *International Journal of e-Collaboration*, *6*(4), 1–11. doi:10.4018/jec.2010100101.

N. Kock (Ed.). (2010b). *Information systems action research: An applied view of emerging concepts and methods*. New York, NY: Springer.

Kock, N. (2011). Using WarpPLS in e-collaboration studies: Descriptive statistics, settings, and key analysis results. *International Journal of e-Collaboration*, *7*(2), 1–17. doi:10.4018/jec.2011040101.

Kock, N. (2011b). Using WarpPLS in e-collaboration studies: Mediating effects, control and second order variables, and algorithm choices. *International Journal of e-Collaboration*, *7*(3), 1–14. doi:10.4018/jec.2011070101.

Kock, N. (2012). *WarpPLS 3.0 User Manual*. Laredo, TX: ScriptWarp Systems.

Kock, N., & Chatelain-Jardón, R. (2011). Four guiding principles for research on evolved information processing traits and technology-mediated task performance. *Journal of the Association for Information Systems*, *12*(10), 684–713.

Rosenthal, R., & Rosnow, R. L. (2007). *Essentials of behavioral research: Methods and data analysis*. Boston, MA: McGraw Hill.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin Company.

Ned Kock is Professor of Information Systems and Director of the Collaborative for International Technology Studies at Texas A&M International University. He holds degrees in electronics engineering (BEE), computer science (MS), and management information systems (PhD). Ned has authored and edited several books, including the bestselling Sage Publications book titled Systems Analysis and Design Fundamentals: A Business Process Redesign Approach*. He has published his research in a number of high-impact journals including* Communications of the ACM, Decision Support Systems, European Journal of Information Systems, European Journal of Operational Research, IEEE Transactions *(various),* Information & Management, Information Systems Journal, Journal of the Association for Information Systems, MIS Quarterly*, and* Organization Science*. He is the Founding Editor-in-Chief of the* International Journal of e-Collaboration*, and Associate Editor of the* Journal of Systems and Information Technology. *His main research interests are biological and cultural influences on human-technology interaction, non-linear structural equation modeling, electronic communication and collaboration, action research, ethical and legal issues in technology research and management, and business process improvement.*