Participant Carelessness: Is It a Substantial Problem With Survey Data?

Shelly Marasi[1], Alison Wall[2] and Kristen Brewer[3]
[1]Tennessee Technological University, College of Business, Department of Decision Sciences & Management, Cookeville, USA
[2]Southern Connecticut State University, School of Business, Department of Management and MIS, New Haven, USA
[3]Eastern Kentucky University, College of Business and Technology, Department of Management, Marketing, and International Business, Richmond, USA
smarasi@tntech.edu
walla4@southernct.edu
kristen.brewer@eku.edu

Abstract: For decades, participant carelessness has been considered a problem in collecting data using surveys. Although participant carelessness cannot be disputed to exist, the impact it has on data quality or the level of influence or bias it produces in results is questionable. The main purpose of this paper is to determine whether participant carelessness is a substantial problem that significantly influences or biases the results of statistical analyses. This is accomplished by analyzing established management relationships through a comparison of the full, careful, and careless samples to determine the impact participant carelessness has on data results regarding correlations, *t*-tests, and simple linear regressions. Four detection approaches were used to identify careless participants individually, in pairs, and in three method combinations. The second purpose of this paper is to use the resampled individual reliability (RIR) approach to detect careless participants and compare it to the individual reliability approach to determine whether the two approaches are fundamentally similar. Data were collected using Mechanical Turk (*N* = 678). Based on the findings, participant carelessness does not appear to be a substantial problem or demonstrate levels of bias in the results in this study. There are two significant differences between the full and careful samples with the *t*-tests and the regression comparisons of fit statistics demonstrate the careful samples to have a weak improvement over the full sample; however, none indicate bias. The findings also suggest that the individual reliability and the RIR approaches are not entirely fundamentally similar.

Keywords: participant carelessness, insufficient effort responding, careless responding, random responding

1. Introduction

Participant carelessness has been argued to be a problem for researchers using surveys to collect data for decades (e.g., Thompson, 1975; Schmitt and Stults, 1985). The challenges of participant carelessness are believed to be growing due to the increasing usage of online data collection. This phenomenon is also referred to as insufficient effort responding or IER (e.g., Huang, et al., 2012; Liu, et al., 2013; Huang, et al., 2015; Huang, Liu and Bowling, 2015; Bowling, et al., 2016; McGonagle, Huang and Walsh, 2016), random responding (e.g., Thompson, 1975; Johnson, 2005; Credé, 2010), and careless responding (Schmitt and Stults, 1985; Meade and Craig, 2012; Maniaci and Rogge, 2014).

Participant carelessness occurs when participants fail to read and/or follow survey instructions or item content or do not take the survey seriously and thereby, may not provide accurate and usable data (Chami-Castaldi, Reynolds and Wallace, 2008; Huang, et al., 2012; Liu, et al., 2013; Bowling, et al., 2016). Therefore, participant carelessness is considered a methodological problem that may lead to measurement error or undesirable effects on the quality and value of the data (Bowling, et al., 2016). This phenomenon has been argued to be an important issue as it may potentially reduce scale reliability (e.g., Huang, et al., 2012; Meade and Craig, 2012) and validity (e.g., Huang, et al., 2012; Aust, et al., 2013; Liu, et al., 2013), lead distinct constructs to be indistinguishable (e.g., Huang, et al., 2012), and cause correlations and other analyses to produce inaccurate results (e.g., Credé, 2010; Maniaci and Rogge, 2014; McGonagle, Huang and Walsh, 2016). For instance, relationships may possibly be altered or obscured between two variables resulting in Type II error (e.g., Meade and Craig, 2012; Maniaci and Rogge, 2014; Huang, Liu and Bowling, 2015; McGonagle, Huang and Walsh, 2016) or Type I error (e.g., Maniaci and Rogge, 2014;

Reference this paper: Marasi, S., Wall, A., and Brewer, K., 2019. Participant Carelessness: Is It a Substantial Problem With Survey Data?. *The Electronic Journal of Business Research Methods,* 17(1), pp. 1-16, available online at www.ejbrm.com

McGonagle, Huang and Walsh, 2016). These bias issues may lead to data being unusable and costly regarding time and survey administration expenses as it decreases the sample size, which may require more data to be collected.

2. Literature Review

As a common methodological problem that possibly produces bias in survey data results, participant carelessness concerns are similar to those associated with common method variance or CMV (McGonagle, Huang and Walsh, 2016) in that while many researchers acknowledge it is a potential problem, it is questioned as to when and how it creates bias in results or reduces the legitimacy of findings (e.g., Spector, 2006). CMV is defined as "variance that is attributable to the measurement method rather than to the construct of interest" (Podsakoff, et al., 2003, p.879). For decades, some researchers have considered CMV to be a major issue in self-report surveys and single source data that needs to be corrected or controlled for when collecting data (e.g., Campbell and Fiske, 1959; Cote and Buckley, 1988; Podsakoff, et al., 2003; Podsakoff, MacKenzie and Podsakoff, 2012). However, others argue that CMV is an overstated problem, a myth, or the bias does not exist to a level that delegitimizes findings (e.g., Spector, 1987; 2006; Vandenberg, 2006; Richardson, Simmering and Sturman, 2009; Fuller, et al., 2016). Since CMV has researchers taking positions on both sides of the spectrum arguing whether it is or is not a problem with data quality and causes bias in the results, participant carelessness should also be examined to determine whether it creates a major issue in data quality and leads to biasing levels in the results. Undoubtedly, no arguments can be made that participant carelessness does not exist (unlike with CMV) as many researchers have experienced careless participants at some point in collecting survey data. Consequently, the argument of participant carelessness not being a serious problem in data analyses deserves examination as has been investigated with CMV.

Therefore, the main purpose of this paper is to determine whether participant carelessness is a substantial problem that significantly influences or biases the results of data analyses using different statistical techniques. This is accomplished by analyzing established management relationships through a comparison of the samples (full, careful, and careless) to determine the impact participant carelessness has on data results regarding correlations, *t*-tests (one-sample *t*-tests and independent samples *t*-tests), and simple linear regression. The analyses are conducted using four participant carelessness detection approaches individually, in pairs, and in combinations of three. The second purpose of this study is to use the resampled individual reliability (RIR) approach as a detection approach and compare it to the individual reliability approach, which have been argued to be similar methods (Curran, 2016) but have yet to be empirically tested according to the authors' knowledge.

3. Theoretical Background

Participant carelessness may occur from participants incorrectly interpreting item content or from being inattentive or careless in responding to the item content. Participant carelessness can occur in various types of surveys, such as those involving academic research, organizational questionnaires (for employees or customers), performance appraisals, and student evaluations.

Participant carelessness may take the form of random responses or nonrandom repeated responses. Random responses entail participants marking responses randomly with no specific pattern. Nonrandom repeated responses involve participants responding in a systematic series or specific sequence, such as straightlining (marking the same response option for every item on a page), near straightlining (straightlining with one item being given a different marking on a page), alternating pattern (marking two or more responses in a rotating pattern), extreme response patterns (marking the extremes responses in a rotating pattern), diagonal pattern in an ascending or descending order, among other patterns.

Participant carelessness may also be unintentional/occasional or intentional. Unintentional or occasional participant carelessness involves participants not fully comprehending some or all item content (which may be due to the wording of the items), having distractions while taking the survey (which may lead participants to be careless in certain parts of the survey or the whole survey), or gradually losing focus over time in completing the survey (and participants may or may not become attentive again). Intentional participant carelessness entails participants purposefully being careless by marking any response, not taking the survey seriously, or speeding through the survey in attempt to complete it as quickly as possible.

3.1 Detection Approaches

Numerous methods have been developed over the years to detect participant carelessness, which make handling this phenomenon easier (Johnson, 2005). Researchers should decide which approach(es) they will utilize to control for participant carelessness before and perhaps during the data collection, even when it is performed in a post hoc manner. Many of the approaches are significantly correlated with one another (e.g., Huang, et al., 2012; Huang, et al., 2015) and demonstrate convergence, suggesting that several approaches together may effectively detect participant carelessness (e.g., Wise and Kong, 2005; Huang, et al., 2015; Bowling, et al., 2016). The decision to utilize a certain approach varies depending on the length of the survey (e.g., short or long), the format of the survey (e.g., online or paper), the practicality and feasibility of the approach's usage, the approach's probability to incorrectly identify attentive participants as careless, and the approach's potential to cause negative reactions in the participants.

Detection approaches are reactive techniques that attempt to control for careless participants after the data has been collected by eliminating them before the analyses (either in a priori or post hoc manner). Therefore, researchers can identify the number of careless participants in a study. A priori detection approaches involve measuring participant carelessness by adopting statements into the survey design before data collection. Post hoc detection approaches involve measuring participant carelessness after data has been collected and generally do not require any special considerations in the survey design. For an overview and description of all detection approaches refer to Huang, et al. (2012) and Curran (2016).

The detection approaches of interest to this study are the instructed response items, the response time, the individual reliability, and the RIR. These five methods were chosen based on the following reasons. First, the instructed response items, the response time, and the individual reliability approaches are three of the five most common detection approaches utilized to identify careless participants (Liu, et al., 2013). The response time and the individual reliability approaches have demonstrated to be powerful techniques in detecting careless participants and valuable in controlling for this phenomenon (Huang, et al., 2012). The instructed response items approach is one of two methods that have shown to result in participants having the greatest positive perceptions towards a survey and its' design with using a detection approach (Huang, et al., 2015). The RIR approach was used since the authors have no knowledge of it being used in a previous study to detect careless participants or compared to the individual reliability approach, which Curran (2016) argues will produce similar results in detecting careless participants.

3.2 Instructed Response Items

This approach was termed by Meade and Craig (2012) and is based on Hough, et al.'s (1990) Nonrandom Response scale. It involves embedding items in a survey that consist of statements that have clear plausible answers. Therefore, participants should provide a specific response given they read the item content. Participants who do not mark the 'correct' response are deemed careless. Item examples include "Please skip this question." and "This is a control question. Mark 'Mostly True' and move on." (Maniaci and Rogge, 2014). These items are interspersed throughout a survey and tend to be placed within the variables' scale items towards the middle or end of a page to better conceal their discovery. Also, these items should not have a uniform wording direction. For instance, the items should not require always marking the fourth response or responses at the lower or higher end of a scale as this may not identify certain careless participants.

There are two main determinations that must be made to use this approach. First, researchers must determine how to eliminate participants based on carelessness. One technique is to use a cutoff score based on an average of the items (e.g., 0 = item incorrectly answered, 1 = item correctly answered) and participants with an average below the predetermined cutoff score, which is determined prior to data collection, are eliminated from the analyses because they are viewed as being careless (e.g., Hough, et al., 1990; Maniaci and Rogge, 2014; Bowling, et al., 2016; McGonagle, Huang and Walsh, 2016). For instance, a study having eight instructed response items with a cutoff score of six will remove participants who have an average less than six (or miss at least three of the instructed response items). The second technique is to eliminate participants for missing one of the instructed response items (e.g., Hauser and Schwarz, 2016). The second determination involves deciding on the number of instructed response items that should be embedded in the survey. Having too few items may not properly identify careless participants as participants may become careless throughout different parts of the survey (or cycle between attentiveness and carelessness). Alternatively, too many items may irritate participants, resulting in participants having negative reactions to the survey or leading them to partake in unpredictable answers for amusement purposes. The recommendation is to incorporate one item per every fifty to one hundred legitimate scale items (Meade and Craig, 2012) or utilize one item on every other page (Maniaci and Rogge, 2014).

3.3 Response Time

This approach was developed by Wise and Kong (2005) and is also referred to as the page time method (Huang, et al., 2012; Bowling, et al., 2016). It analyzes the entire time spent on completing the survey or a webpage. The assumption of this approach is that extremely short response times indicate participant carelessness since a minimum amount of time is needed to complete a survey as some degree of time for cognitive processing is needed to read, understand, and then respond to each item (Huang, et al., 2012; Meade and Craig, 2012; Maniaci and Rogge, 2014; Huang, et al., 2015; Bowling, et al., 2016). For example, a participant that completes a survey consisting of fifty items in one minute would demonstrate carelessness as all items could not have been read, comprehended, and accurately answered within the short amount of time.

This approach can only be used with online surveys and a cutoff time must be identified. There are two ways a cutoff time can be established. First, an average response time for completing the survey (e.g., Weathers and Bardakci 2015) or webpage (Huang, et al., 2012) can be calculated and participants who fall significantly below the average are considered careless. For example, when the average time for survey completion is seven minutes, a participant who finishes it in two minutes is deemed a careless participant. Second, a response time per item can be established and summated prior to data collection and participants who do not meet the overall cutoff time are eliminated for carelessness. The recommendation is a cutoff time of two seconds per item (e.g., Bowling, et al., 2016). However, long response time outliers (which may be due to participants taking a break or being distracted while completing the survey) need to be accounted for in the calculating the average response time.

3.4 Individual Reliability

This approach was created by Jackson (1977) and is also referred to as the “even-odd consistency” approach (e.g., Meade and Craig, 2012; Maniaci and Rogge, 2014). It involves dividing a variable’s scale items using an even- and odd-split, creating half-scale scores or two subscales of an overall variable scale (an even and an odd subscale). The split is determined based on the order the items appear in the survey. For example, a six-item scale would have items appearing first, third, and fifth in the survey being in the odd subscale and the even subscale consisting of items appearing second, fourth, and sixth. Negatively worded items are reverse-coded beforehand. The two subscales are then compared for within-person reliability through correlations. Therefore, comparison correlations are computed for every variable scale per each participant. This approach is based on the foundation that items belonging to the same scale are expected to correlate with each other and it is suggested that correlations less than .30 indicate careless participants (Jackson, 1977).

This approach requires variable scales that have enough items to form the two subscales as one item subscales are unusable (Curran, 2016). Therefore, variable scales must consist of at least four items for this approach to be used properly. The recommendation is for a minimum of six items per variable scale since the subscale scores are constrained by the number of items in the scale. This approach can be used with unidimensional scales and multi-dimensional scales, given there are enough items in each subdimension to create two subscales (Curran, 2016).

3.5 RIR

This approach is proposed as an alternative to the individual reliability approach, but the division of items for the two subscales is based on randomness (Curran, 2016). For example, a six-item scale may be divided by items 1, 4, and 5 being in one subscale and the second subscale consisting of items 2, 3, and 6. The rationale for this approach is that since there is nothing fundamentally unique about the subscales' composition following the even-odd-split, similar scores should be produced from randomly drawn subscales (Curran, 2016). Additionally, this approach allows for multiple pairs of subscales to be created with the assumption that none of the pairs are better than the other, including the even-odd-split subscales (Curran, 2016). Even though multiple random assignments for a scale can be created, only a single random assignment of a variable scale's items per participant is necessary. Similar to the individual reliability approach, this approach has the same requirements for the number of scale items (minimum of four items per scale) and recommendation of a correlation cutoff score of .30.

4. Research Questions

The main objective of this study is to determine whether participant carelessness is a substantial problem for researchers that significantly influences or biases results from different statistical analyses. This is determined by identifying whether there are significant differences in correlations, *t*-tests, and simple linear regression, regarding the inclusion and exclusion of careless participants in a sample involving established management relationships. Specifically, the constructs included job satisfaction (JS), organizational commitment (OC), and organizational citizenship behaviors (OCB). Meta-analyses demonstrate that higher levels of JS and OC result in greater engagement of OCB and that JS and OC are correlates (e.g., LePine, Erez and Johnson, 2002; Meyer, et al., 2002). Since these relationships have been frequently researched and recognized in the management literature, differences in the analyses between the samples that include and exclude careless participants should be evident. Therefore, the following research questions are proposed:

> *Research Question 1: To what extent does participant carelessness influence or bias the results of different statistical analyses?*
> *Research Question 2: To what extent are the individual reliability and the RIR approaches fundamentally similar?*

5. Method

Participants were recruited using an online survey organization, Mechanical Turk. The compensation for participants was twenty-five cents. Participants consented to participate in the survey and then were given instructions to complete it. Participants were anonymous to the researchers. A response rate is unable to be identified due to the operation of Mechanical Turk.

The sample consisted of 678 respondents residing in the U.S. Participants ranged from 18 to 72 years old with the mean age being 35 years. The sample was predominately comprised of females (56%), whites (78%), and those possessing a bachelor's degree or higher (57%).

5.1 Measures

JS was assessed with Cammann, et al.'s (1983) three-item Job Satisfaction scale, which was measured using a 7-point Likert-type scale (1 = strongly disagree, 7 = strongly agree). One item was reverse-coded. An item example is "All in all, I am satisfied with my job." The items were averaged to produce the scale (α = .93 full sample).

OC was assessed with Mowday, Steers and Porter's (1979) short-version Organizational Commitment scale consisting of eight items (Commeiras and Fournier, 2001), which was measured using a 7-point Likert-type scale (1 = strongly disagree, 7 = strongly agree). "I really care about the fate of this organization" is an item example. The items were averaged to produce the scale (α = .93 full sample).

OCB were measured using Williams and Anderson's (1991) fourteen-item Organizational Citizenship Behaviors scale, which was measured using a 5-point Likert-type scale (1 = strongly disagree, 5 = strongly agree). Three items were reverse-coded. An item example is "Helps others who have been absent." The items were averaged to produce the scale (α = .83 full sample).

Three additional scales and five demographic questions were included in the survey to produce a medium length survey and receive a better representation of participant carelessness. The following scales were included:

Williams and Anderson's (1991) In-role Behavior scale, which consists of seven items with two being reverse-coded; Burton, et al.'s (1998) Private Label Attitude scale, consisting of the five positively-worded items; Miller and Chiodo's (2008) Attitudes towards the Color Blue scale, including the four positively-worded items. All three additional scales were measured using a 5-point Likert-type scale (1 = strongly disagree, 5 = strongly agree). Therefore, there were fifty questions in the survey with an expected completion time of ten to fifteen minutes depending on the participant.

5.2 Approaches

Three instructed response items were used with one being inputted on every other webpage (Maniaci and Rogge, 2014). An item example is "Select disagree for this item." The cutoff score was one. Therefore, participants who missed any of the instructed response items were deemed careless.
The response time approach was utilized with a cutoff time set at one-third of the average time of the survey completion after outliers (over twenty-five minutes) were removed. The average time for survey completion was 12 minutes and 43 seconds, yielding a cutoff time of 3 minutes and 11 seconds. Participants that completed the survey in less time than the cutoff time were deemed careless.

The individual reliability and the RIR approaches were used with a cutoff score of .30 and participants with correlations less than .30 were identified as careless. Therefore, participants that did not meet the cutoff on either the OC or OCB scale were deemed careless. For the RIR approach, the first subscale of OC consisted of items 1, 2, 5, and 6 and the second subscale included items 3, 4, 7, and 8; whereas, the first OCB subscale included items 1, 4, 6, 8, 9, 11, and 13 and the second subscale consisted of items 2, 3, 5, 7, 10, 12, and 14. Both of these approaches were unable to be used on the JS scale since it is does not meet the required four-item threshold.

6. Results

Three samples were yielded from the approaches (individually and in combination): full, careful, and careless. Following suggestions from scholars (e.g., McGonagle, Huang and Walsh, 2016), analyses included a comparison of the full sample to the subsamples (careful and careless) and a comparison of the subsamples.

Correlations were compared between the three samples by conducting the Fisher's z-transformation to calculate a *z* score using VassarStats (Lowry, 2018). One-sample *t*-tests were conducted in SPSS 25 to compare the full sample's mean to the careful and careless samples' means using the full sample's mean as the population (or test) mean. Independent samples *t*-tests were conducted in SPSS 25 to compare the means of the careful sample to the careless sample. Cohen's *d* was conducted in SPSS 25 to determine the effect sizes of the significant *t*-tests. Simple linear regressions were conducted in SPSS 25 for analyzing the JS-OCB and OC-OCB relationships. The independent variables were mean centered for better result interpretation (Cohen, et al., 2003). The Chow test was conducted in SPSS 25 to determine whether the linear regressions are equal across the careful and careless samples. An examination of the fit statistics (R^2, adjusted R^2, and standard error of the estimate or SEE) from the regression produced from SPSS 25 were used to determine whether there was a difference in the linear models between the full sample and the careful and careless samples (Hair Jr., et al., 2010). The differences between the regression models for the samples were determined by the researchers as being minimal (a miniscule difference within .02 in the R^2, adjusted R^2, and/or SEE), small (a slight difference between .03 and .09 in the R^2, adjusted R^2, and/or SEE), moderate (a difference between .10 and .20 in the R^2, adjusted R^2, and/or SEE that demonstrates bias), or large differences (a huge difference in the R^2, adjusted R^2, and/or SEE and/or alters the direction and significance of the beta coefficient), with significance being considered for moderate and large differences. The full sample's fit statistics for the JS-OCB relationship are R^2=.11, adjusted R^2=.11, SEE=.51, β = .326**, and for the OC-OCB relationship are R^2=.15, adjusted R^2=.15, SEE=.50, β=.386**.

Each approach was examined individually and in combinations. For the two approach combinations, all possible combinations were evaluated except for the combination of the individual reliability and the RIR approaches since they are argued to be similar approaches. The three approach combination involved the instructed response items, the response time, and either the individual reliability or the RIR approaches. Therefore, there were four individual, five paired combinations, and two combinations of three approaches utilized for the statistical analyses (refer to the Appendix for analyses results).

One participant (.15%) was identified as careless in all four approaches. Thirty-seven participants (5.46%) were deemed careless by any combination of three approaches, while 184 participants (27.14%) were determined to be careless by any combination of two approaches. Additionally, the individual reliability and the RIR approaches identified 162 of the same participants as careless, producing a 70.5% overlap.

7. Discussion

Participant carelessness does occur; however, its influence on data quality and statistical analyses may not be a major issue as argued by some researchers. This paper examines whether participant carelessness is a substantial problem and has a significant influence or bias on results. According to the researcher's knowledge, this is the first paper to utilize the RIR approach to detect careless participant and compare it to the individual reliability approach for fundamental similarities. The findings of this study offer several important inferences.

Research Question 1 addresses the extent to which participant carelessness influences or biases the results of different statistical analyses. The findings of the correlations, *t*-tests, and regressions between the samples from the different detection approaches (individually and in combination) provide implications for this research question. However, comparisons between the full and careful samples are most important for making inferences since the elimination of the careless participants results in the careful sample being used for statistical analyses rather than the full sample.

For the correlation analyses, most of the significant differences are between the careful and careless samples with a few being between the full and careless samples. However, there are no significant differences between the full and careful sample in the correlation analyses. Most of the significant differences in the *t*-tests are between the careful and careless samples and between the full and careless samples. However, many of the *t*-tests significant differences have a weak effect size. There are only two significant differences in the *t*-tests between the full and careful samples that show the careful sample has a higher mean than the full sample; however, both have a weak effect size and involve the RIR approach. For the simple linear regression comparisons, there are many significant differences between the careful and careless samples. There are also several moderate to large differences between the full and careless samples in the regression comparisons. There are only minimal or small differences in the regression comparisons between the full and careful samples, which do not appear to demonstrate bias. However, most of the differences between the full and careful samples show the careful samples' regression models were negligibly or slightly better than the full samples' regression models.

Therefore, most of the significant differences between the full and careful samples indicate the careful samples have a weak increase in the means and slight improvement in the regression fit statistics (and beta coefficients). This demonstrates that in some instances the results of the full sample (and inclusion of careless participants) are slightly deflated, while other results are slightly inflated. However, these significant differences did not demonstrate the full samples' results to be altered to an extent that causes them from being misinterpreted or delegitimized, such as eliminating or extremely changing significant relationships. Thus, the findings of this study suggest participant carelessness may have negligible or little overall impact in analyses results and may not create a severe issue in data quality by highly influencing or biasing the results of statistical analyses.

This study's implication of participant carelessness contradicts other scholars' claims and findings (e.g., McGonagle, Huang and Walsh, 2016; DeSimone, et al., 2018). However, the difference in findings from this study and others may be due to multiple factors. For instance, the data in this study is real (rather than fully or partially simulated), established management relationships were examined (rather than a single scale or different relationships), and the data was not altered to force a specific number of participants to be careless at specific levels. Thus, participant carelessness may exist, but this study's results suggest that it is not to an extent that delegitimizes the findings or reduces data quality, which is similar to one of the arguments regarding CMV.

However, it should be noted that although this study did not find that the significant differences delegitimized the results or reduced data quality, this may not be the case in other studies using real data.

Research Question 2 involves identifying the extent to which the individual reliability and the RIR approaches are fundamentally similar. The approaches identified many of the same careless participants (n = 162), producing a 70.5% overlap. However, the significant differences detected in the analyses varied between the approaches. In fact, there were thirteen different significant differences identified in the analyses when the approaches were compared individually or in combinations. Additionally, both significant differences in the t-tests between the full and careful samples involve the RIR approach in combination with other approaches, while the individual reliability approach in the same combinations are not significant. Therefore, the results of this study demonstrate the RIR approach does not detect the same participants as careless or perform the same. Thus, the RIR and individual reliability approaches are not entirely fundamentally similar according to the results of this study.

7.1 Research Implications

Participant carelessness has been a major concern for online data collection. However, the results of this study demonstrate that it may not be a major issue in data quality or creating bias in results. Although participant carelessness was not a substantial concern in this study, it may be in other studies. Therefore, the best technique to ensure it does not become a major problem is to utilize at least one detection approach in an online data collection since participants will vary and the results may be different in other studies. Additionally, the RIR and individual reliability approaches appear to not be entirely fundamentally similar and interchangeable. However, the RIR approach may still be a good detection approach for participant carelessness.

Individually, the individual reliability and the RIR approaches identified the highest levels of participant carelessness. These results support previous research that the individual reliability approach is very effective as it outperforms other methods in determining careless participants (e.g., Huang, et al., 2012, Meade and Craig, 2012). The response time approach alone was not very successful in detecting careless participants in this study as it only identified five participants as careless. This finding contradicts previous research that found the response time approach to be a reliable (Wise and Kong, 2005) and effective detection approach (e.g., Huang, et al., 2012; McGonagle, Huang and Walsh, 2016). The instructed response items approach found a modest amount of careless participants and appeared to detect careless participants at different phases of the survey (e.g., beginning, middle, and end).

7.2 Limitations and Future Directions

The first limitation is that this study only included a few scales that have an established relationship and therefore, direct evidence that the results will be similar with other relationships (established or not) or scales cannot be provided. Additionally, this study used online survey data collection methods, which may lead results to not be duplicated with other survey methodologies (e.g., paper surveys). Thus, a future path is to compare the extent of careless participants across different survey methodologies.

This study only utilizing four detection approaches is another limitation. Other detection approaches were not examined and may produce different results than found in this study. Therefore, a future avenue is to explore the influence other detection approaches have on the same statistical analyses. Additionally, another limitation may involve the detection approaches providing false positives (Aust, et al., 2013), which may have occurred with the individual reliability and the RIR approaches since they both identified a large number of careless participants but there was not a complete overlap between the two approaches.

CMV may be a potential limitation. However, two procedural remedies were used, including altering the item order and providing anonymity to participants (Podsakoff, et al., 2003). Harmon's single-factor test was also conducted, which showed that the items did not load on one factor and one factor did not account for most of the covariance (Podsakoff, et al., 2003). Therefore, CMV is unlikely to be present or exist at levels that bias the results.

The external validity or generalizability of the results is the final limitation of this study. The findings may not be generalizable to countries other than the U.S. since the sample was comprised of only U.S. residents. Therefore, a future avenue may be to replicate this study with participants from other countries to identify whether the findings are similar or different.

A final future path is to further investigate the fundamental similarities between the individual reliability and the RIR approaches to identify whether they are interchangeable and produce similar results or replicate and substantiate the findings of this study.

References

Aust, F., Diedenhofen, B., Ullrich, S. & Musch, J., 2013. Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods,* 45(2), pp. 527-535.

Bowling, N. A. et al., 2016. Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology,* 111(2), pp. 218-229.

Burton, S., Lichtenstein, D. R., Netemeyer, R. G. & Garretson, J. A., 1998. A scale for measuring attitude toward private label products and an examination of its psychological and behavioral correlates. *Journal of the Academy of Marketing Science,* 26(4), pp. 293-306.

Cammann, C., Fichman, M., Jenkins, D. & Klesh, J. R., 1983. Assessing the attitudes and perceptions of organizational members. In: S. E. Seashore, E. E. Lawler, P. H. Mirvis & C. Cammann, eds. *Assessing organizational change: A guide to methods, measures, and practices.* New York: John Wiley, pp. 71-138.

Campbell, D. T. & Fiske, D. W., 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin,* 56(2), pp. 81-105.

Chami-Castaldi, E., Reynolds, N. & Wallace, J., 2008. Individualised rating-scale procedure: A means of reducing response style contamination in survey data?. *The Electronic Journal of Business Research Methods,* 6(1), pp. 9-20.

Cohen, J., Cohen, P., West, S. G. & Aiken, L. S., 2003. *Applied multiple regression/correlation analysis for the behavioral sciences.* 3rd ed. Mahwah(New Jersey): Lawrence Erlbaum Associates, Inc..

Commeiras, N. & Fournier, C., 2001. Critical evaluation of Porter et al.'s Organizational Commitment Questionnaire: Implications for researchers. *The Journal of Personal Selling & Sales Management,* 21(3), pp. 239-245.

Cote, J. A. & Buckley, M. R., 1988. Measurement error and theory testing in consumer research: An illustration of the importance of construct validation. *Journal of Consumer Research,* 14(4), pp. 579-582.

Credé, M., 2010. Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement,* 70(4), pp. 596-612.

Curran, P. G., 2016. Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology,* 66(1), pp. 4-19.

DeSimone, J. A., DeSimone, A. J., Harms, P. D. & Wood, D., 2018. The differential impacts of two forms of insufficient effort responding. *Applied Psychology: An International Review,* 67(2), pp. 309-338.

Fuller, C. M. et al., 2016. Common method variance detection in business research. *Journal of Business Research,* 69(8), pp. 3192-3198.

Hair Jr., J. F., Black, W. C., Babin, B. J. & Anderson, R. E., 2010. *Multivariate data analysis.* 7th ed. Upper Saddle River(NJ): Prentice Hall.

Hauser, D. J. & Schwarz, N., 2016. Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods,* 48(1), pp. 400-407.

Hough, L. M. et al., 1990. Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology,* 75(5), pp. 581-595.

Huang, J. L., Bowling, N. A., Liu, M. & Li, Y., 2015a. Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology,* 30(2), pp. 299-311.

Huang, J. L. et al., 2012. Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology,* 27(1), pp. 99-114.

Huang, J. L., Liu, M. & Bowling, N. A., 2015b. Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology,* 100(3), pp. 828-845.

Jackson, D. N., 1977. *Jackson vocational interest survey manual.* Port Huron(MI): Research Psyhologists Press.

Johnson, J. A., 2005. Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality,* 39(1), pp. 103-129.

LePine, J. A., Erez, A. & Johnson, D. E., 2002. The nature and dimensionality of organizational citiznenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology,* 87(1), pp. 52-65.

Liu, M., Bowling, N. A., Huang, J. L. & Kent, T., 2013. Insufficient effort responding to surveys as a threat to validity: The perceptions and practices of SIOP members. *The Industrial-Organizational Psychologist,* 51(1), pp. 32-38.

Lowry, R., 2018. *Significance of the difference between two correlation coefficients.* [Online] Available at: http://vassarstats.net/rdiff.html

Maniaci, M. R. & Rogge, R. D., 2014. Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality,* Volume 48, pp. 61-83.

McGonagle, A. A., Huang, J. L. & Walsh, B. M., 2016. Insufficient effort survey responding: An under-appreciated problem in work and organisational helath psychology research. *Applied Psychology: An International Review,* 65(2), pp. 287-321.

Meade, A. W. & Craig, S. B., 2012. Identifying careless responses in survey data. *Psychological Methods,* 17(3), pp. 437-455.

Meyer, J. P., Stanley, D. J., Herscovitch, L. & Topolnytsky, L., 2002. Affective, continuance, and normative commitment to the organization: A meta-analysis of antecedents, correlates, and consequences. *Journal of Vocational Behavior,* 61(1), pp. 20-52.

Miller, B. & Chiodo, B., 2008, October. *Academic entitlement: Adapting the equity preference questionnaire for a university setting.* s.l.:Paper presented at the meeting of Southern Management Association, St. Pete Beach, FL..

Mowday, R. T., Steers, R. M. & Porter, L. W., 1979. The measurement of organizational commitment. *Journal of Vocational Behavior,* 14(2), pp. 224-247.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. & Podsakoff, N. P., 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology,* 88(5), pp. 879-903.

Podsakoff, P. M., MacKenzie, S. B. & Podsakoff, N. P., 2012. Sources of method bias in social science research and recommendations on how to control it. *Annual Review Psychology,* 63(1), pp. 539-569.

Richardson, H. A., Simmering, M. J. & Sturman, M. C., 2009. A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods,* 12(4), pp. 762-800.

Schmitt, N. & Stults, D. M., 1985. Factors defined by negatively keyed items: The result of careless respondents?. *Applied Psychologial Measurement,* 9(4), pp. 367-373.

Spector, P. E., 1987. Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem?. *Journal of Applied Psychology,* 72(3), pp. 438-443.

Spector, P. E., 2006. Method variance in organizational research: Truth or urban legend?. *Organizational Research Methods,* 9(2), pp. 221-232.

Thompson, A. H., 1975. Random responding and the questionnaire measurement of psychoticism. *Social Behavior and Personality: An International Journal,* 3(2), pp. 111-115.

Vandenberg, R. J., 2006. Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea?. *Organizational Research Methods,* 9(2), pp. 194-201.

Weathers, D. & Bardakci, A., 2015. Can response variance effectively identify careless respondents to multi-item, unidimensional scales?. *Journal of Marketing Analytics,* 3(2), pp. 96-107.

Williams, L. J. & Anderson, S. E., 1991. Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management,* 17(3), pp. 601-617.

Wise, S. L. & Kong, X., 2005. Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education,* 18(2), pp. 163-183.

Appendix

Analysis of Results

	Sample sizes	Cronbach's alpha reliabilities	
		Careful	Careless
Instructed Response Items	Careful=565 (83.3%) Careless=113 (16.7%)	JS: α = .94 OC: α = .93 OCB: α = .85	JS: α = .88 OC: α = .90 OCB: α = .86
Response Time	Careful=673 (99.3%) Careless=5 (0.7%)	JS: α = .93 OC: α = .93 OCB: α = .85	JS: α = .51 OC: α = .89 OCB: α = .85
Individual Reliability	Careful=449 (66.2%) Careless=229 (33.8%)	JS: α = .93 OC: α = .94 OCB: α = .86	JS: α = .94 OC: α = .90 OCB: α = .83
RIR	Careful=448 (66.1%) Careless=230 (33.9%)	JS: α = .94 OC: α = .94 OCB: α = .87	JS: α = .92 OC: α = .89 OCB: α = .81
Instructed Response Items and Individual Reliability	Careful=366 (54%) Careless=312 (46%)	JS: α = .94 OC: α = .94 OCB: α = .86	JS: α = .92 OC: α = .91 OCB: α = .84
Instructed Response Items and RIR	Careful=378 (55.8%) Careless=300 (44.2%)	JS: α = .95 OC: α = .94 OCB: α = .87	JS: α = .91 OC: α = .90 OCB: α = .82
Response Time and Individual Reliability	Careful=446 (65.8%) Careless=232 (34.2%)	JS: α = .93 OC: α = .93 OCB: α = .86	JS: α = .93 OC: α = .91 OCB: α = .84
Response Time and RIR	Careful=445 (65.6%) Careless=233 (34.4%)	JS: α = .94 OC: α = .94 OCB: α = .86	JS: α = .92 OC: α = .89 OCB: α = .82
Instructed Response Items and Response Time	Careful=563 (83%) Careless=115 (17%)	JS: α = .94 OC: α = .93 OCB: α = .84	JS: α = .88 OC: α = .91 OCB: α = .87
Instructed Response Items, Response Time, and Individual Reliability	Careful=364 (53.7%) Careless=314 (46.3%)	JS: α = .94 OC: α = .94 OCB: α = .86	JS: α = .92 OC: α = .91 OCB: α = .84
Instructed Response Items, Response Time, and RIR	Careful=376 (55.5%) Careless=302 (44.5%)	JS: α = .95 OC: α = .94 OCB: α = .86	JS: α = .91 OC: α = .90 OCB: α = .83

	Correlations		
	Full/Careful	Full/Careless	Careful/Careless
Instructed Response Items	JS-OC: $.763^{**}/.790^{**}$; $z=-1.19, p<.12$ JS-OCB: $.326^{**}/.316^{**}$; $z=.20, p<.43$ OC-OCB: $.386^{**}/.417^{**}$; $z=-.65, p<.26$	**JS-OC: $.763^{**}/.574^{**}$; $z=3.40, p<.01$** JS-OCB: $.326^{**}/.432^{**}$; $z=-1.21, p<.12$ OC-OCB: $.386^{**}/.353^{**}$; $z=.37, p<.36$	**JS-OC: $.790^{**}/.574^{**}$; $z=4.01, p<.01$** JS-OCB: $.316^{**}/.432^{**}$; $z=-1.30, p<.10$ OC-OCB: $.417^{**}/.353^{**}$; $z=.72, p<.24$
Response Time	JS-OC: $.763^{**}/.764^{**}$; $z=-.04, p<.49$ JS-OCB: $.326^{**}/.323^{**}$; $z=.06, p<.48$ OC-OCB: $.386^{**}/.377^{**}$; $z=.19, p<.43$	JS-OC: $.763^{**}/.433$; $z=.76, p<.23$ JS-OCB: $.326^{**}/-.097$; $z=.62, p<.27$ OC-OCB: $.386^{**}/.558$ $z=-.31, p<.38$	JS-OC: $.764^{**}/.433$; $z=.77, p<.23$ JS-OCB: $.323^{**}/-.097$; $z=.61, p<.28$ OC-OCB: $.377^{**}/.558$; $z=-.33, p<.38$
Individual Reliability	JS-OC: $.763^{**}/.778^{**}$; $z=-.61, p<.28$ JS-OCB: $.326^{**}/.389^{**}$; $z=-1.18, p<.12$ OC-OCB: $.386^{**}/.432^{**}$; $z=-.91, p<.19$	JS-OC: $.763^{**}/.737^{**}$; $z=.77, p<.23$ JS-OCB: $.326^{**}/.216^{**}$; $z=1.55, p<.07$ OC-OCB: $.386^{**}/.295^{**}$; $z=1.34, p<.10$	JS-OC: $.778^{**}/.737^{**}$; $z=1.18, p<.12$ **JS-OCB: $.389^{**}/.216^{**}$; $z=2.34, p<.01$** **OC-OCB: $.432^{**}/.295^{**}$; $z=1.94, p<.03$**
RIR	JS-OC: $.763^{**}/.785^{**}$; $z=-.90, p<.19$ JS-OCB: $.326^{**}/.373^{**}$; $z=-.88, p<.19$ OC-OCB: $.386^{**}/.436^{**}$; $z=-.99, p<.17$	JS-OC: $.763^{**}/.715^{**}$; $z=1.38, p<.09$ JS-OCB: $.326^{**}/.218^{**}$; $z=1.52, p<.07$ **OC-OCB: $.386^{**}/.257^{**}$ $z=1.88, p<.04$**	**JS-OC: $.785^{**}/.715^{**}$; $z=1.97, p<.03$** **JS-OCB: $.373^{**}/.218^{**}$; $z=2.09, p<.02$** **OC-OCB: $.436^{**}/.257^{**}$; $z=2.51, p<.01$**
Instructed Response Items and Individual Reliability	JS-OC: $.763^{**}/.795^{**}$; $z=-1.25, p<.11$ JS-OCB: $.326^{**}/.392^{**}$; $z=-1.16, p<.13$ OC-OCB: $.386^{**}/.455^{**}$; $z=-1.29, p<.10$	JS-OC: $.763^{**}/.726^{**}$; $z=1.21, p<.12$ JS-OCB: $.326^{**}/.244^{**}$; $z=1.30, p<.10$ OC-OCB: $.386^{**}/.302^{**}$; $z=1.39, p<.09$	**JS-OC: $.795^{**}/.726^{**}$; $z=2.13, p<.02$** **JS-OCB: $.392^{**}/.244^{**}$; $z=2.13, p<.02$** **OC-OCB: $.455^{**}/.302^{**}$; $z=2.32, p<.02$**
Instructed Response Items and RIR	JS-OC: $.763^{**}/.797^{**}$; $z=-1.35, p<.09$ JS-OCB: $.326^{**}/.366^{**}$; $z=-.71, p<.24$ OC-OCB: $.386^{**}/.448^{**}$; $z=-1.17, p<.13$	JS-OC: $.763^{**}/.716^{**}$; $z=1.49, p<.07$ JS-OCB: $.326^{**}/.262^{**}$; $z=1.01, p<.16$ OC-OCB: $.386^{**}/.293^{**}$; $z=1.51, p<.07$	**JS-OC: $.797^{**}/.716^{**}$; $z=2.46, p<.01$** JS-OCB: $.366^{**}/.262^{**}$; $z=1.49, p<.07$ **OC-OCB: $.448^{**}/.293^{**}$; $z=2.32, p<.02$**
Response Time and Individual Reliability	JS-OC: $.763^{**}/.778^{**}$; $z=-.60, p<.28$ JS-OCB: $.326^{**}/.386^{**}$; $z=-1.12, p<.14$ OC-OCB: $.386^{**}/.418^{**}$; $z=-.62, p<.27$	JS-OC: $.763^{**}/.734^{**}$; $z=.86, p<.20$ JS-OCB: $.326^{**}/.224^{**}$; $z=1.44, p<.08$ OC-OCB: $.386^{**}/.327^{**}$; $z=.88, p<.19$	JS-OC: $.778^{**}/.734^{**}$; $z=1.26, p<.11$ **JS-OCB: $.386^{**}/.224^{**}$; $z=2.20, p<.02$** OC-OCB: $.418^{**}/.327^{**}$; $z=1.30, p<.10$
Response Time and RIR	JS-OC: $.763^{**}/.786^{**}$; $z=-.94, p<.18$ JS-OCB: $.326^{**}/.369^{**}$; $z=-.80, p<.22$ OC-OCB: $.386^{**}/.422^{**}$; $z=-.70, p<.25$	JS-OC: $.763^{**}/.712^{**}$; $z=1.47, p<.08$ JS-OCB: $.326^{**}/.227^{**}$; $z=1.41, p<.08$ OC-OCB: $.386^{**}/.293^{**}$; $z=1.38, p<.09$	**JS-OC: $.786^{**}/.712^{**}$; $z=2.09, p<.02$** **JS-OCB: $.369^{**}/.227^{**}$; $z=1.92, p<.03$** **OC-OCB: $.422^{**}/.293^{**}$; $z=1.82, p<.04$**
Instructed Response Items and Response Time	JS-OC: $.763^{**}/.792^{**}$; $z=-1.28, p<.11$ JS-OCB: $.326^{**}/.314^{**}$; $z=.23, p<.41$ OC-OCB: $.386^{**}/.407^{**}$; $z=-.44, p<.33$	**JS-OC: $.763^{**}/.575^{**}$; $z=3.41, p<.01$** JS-OCB: $.326^{**}/.443^{**}$; $z=-1.35, p<.09$ OC-OCB: $.386^{**}/.417^{**}$; $z=-.36, p<.36$	**JS-OC: $.792^{**}/.575^{**}$; $z=4.08, p<.01$** JS-OCB: $.314^{**}/.443^{**}$; $z=-1.46, p<.08$ OC-OCB: $.407^{**}/.417^{**}$; $z=-.12, p<.46$
Instructed Response Items, Response Time, and Individual Reliability	JS-OC: $.763^{**}/.797^{**}$; $z=-1.33, p<.10$ JS-OCB: $.326^{**}/.390^{**}$; $z=-1.13, p<.13$ OC-OCB: $.386^{**}/.441^{**}$; $z=-1.02, p<.16$	JS-OC: $.763^{**}/.723^{**}$; $z=1.31, p<.10$ JS-OCB: $.326^{**}/.249^{**}$; $z=1.23, p<.11$ OC-OCB: $.386^{**}/.325^{**}$; $z=1.02, p<.16$	**JS-OC: $.797^{**}/.723^{**}$; $z=2.28, p<.02$** **JS-OCB: $.390^{**}/.249^{**}$; $z=2.04, p<.03$** **OC-OCB: $.441^{**}/.325^{**}$; $z=1.76, p<.04$**
Instructed Response Items, Response Time, and RIR	JS-OC: $.763^{**}/.798^{**}$; $z=-1.39, p<.09$ JS-OCB: $.326^{**}/.363^{**}$; $z=-.65, p<.26$ OC-OCB: $.386^{**}/.435^{**}$; $z=-.91, p<.19$	JS-OC: $.763^{**}/.713^{**}$; $z=1.58, p<.06$ JS-OCB: $.326^{**}/.267^{**}$; $z=.93, p<.18$ OC-OCB: $.386^{**}/.318^{**}$; $z=1.12, p<.14$	**JS-OC: $.798^{**}/.713^{**}$; $z=2.57, p<.01$** JS-OCB: $.363^{**}/.267^{**}$; $z=1.37, p<.09$ **OC-OCB: $.435^{**}/.318^{**}$; $z=1.76, p<.04$**

Note: Significant findings are in boldface. $^{*}p<.05$. $^{**}p<.01$.

	Independent samples *t*-tests (Careful/Careless)
Instructed Response Items	JS: $t(183) = -.44$, $p = .663$[a] **OC: $t(190) = -2.76$, $p = .006$[a]; Cohen's $d = .26$ (small)** **OCB: $t(676) = 3.83$, $p < .001$; Cohen's $d = .38$ (small)**
Response Time	JS: $t(676) = 1.49$, $p = .138$ **OC: $t(676) = 2.38$, $p = .017$; Cohen's $d = 1.08$ (large)** **OCB: $t(676) = 4.64$, $p < .001$; Cohen's $d = 1.98$ (huge)**
Individual Reliability	**JS: $t\ (676) = 2.16$, $p = .031$; Cohen's $d = .17$ (very small)** OC: $t\ (676) = 1.60$, $p = .11$ OCB: $t\ (676) = -.15$, $p = .878$
RIR	**JS: $t\ (676) = 2.02$, $p = .044$; Cohen's $d = .16$ (very small)** **OC: $t\ (512) = 2.15$, $p = .032$[a]; Cohen's $d = .17$ (very small)** OCB: $t\ (676) = 1.59$, $p = .113$
Instructed Response Items and Individual Reliability	JS: $t\ (676) = 1.41$, $p = .159$ OC: $t\ (673) = -.03$, $p = .978$[a] OCB: $t\ (676) = 1.83$, $p = .068$
Instructed Response Items and RIR	JS: $t\ (676) = 1.32$, $p = .188$ OC: $t\ (668) = .49$, $p = .623$[a] **OCB: $t\ (676) = 3.28$, $p = .001$; Cohen's $d = .25$ (small)**
Response Time and Individual Reliability	**JS: $t\ (676) = 2.37$, $p = .018$; Cohen's $d = .19$ (small)** **OC: $t\ (676) = 1.99$, $p = .048$; Cohen's $d = .16$ (very small)** OCB: $t\ (676) = .42$, $p = .675$
Response Time and RIR	**JS: $t\ (676) = 2.23$, $p = .026$; Cohen's $d = .18$ (very small)** **OC: $t\ (676) = 2.46$, $p = .014$; Cohen's $d = .20$ (small)** **OCB: $t\ (676) = 2.16$, $p = .031$; Cohen's $d = .18$ (very small)**
Instructed Response Items and Response Time	JS: $t\ (188) = -.27$, $p = .79$[a] **OC: $t\ (186) = -2.19$, $p = .03$[a]; Cohen's $d = .21$ (small)** **OCB: $t\ (151) = 4.07$, $p < .001$[a]; Cohen's $d = .44$ (small)**
Instructed Response Items, Response Time, and Individual Reliability	JS: $t\ (676) = 1.52$, $p = .128$ OC: $t\ (673) = .27$, $p = .787$[a] **OCB: $t\ (676) = 2.32$, $p = .021$; Cohen's $d = .18$ (very small)**
Instructed Response Items, Response Time, and RIR	JS: $t\ (676) = 1.43$, $p = .154$ OC: $t\ (667) = .79$, $p = .429$[a] **OCB: $t\ (676) = 3.77$, $p < .001$; Cohen's $d = .29$ (small)**

Note: Significant findings are in boldface.
[a]Equal variances not assumed.

	One-sample *t*-tests	
	Full/Careful	Full/Careless
Instructed Response Items	JS: $t(564) = -.15$, $p = .878$ OC: $t(564) = -.94$, $p = .35$ OCB: $t(564) = 1.58$, $p = .114$	JS: $t(112) = .41$, $p = .681$ **OC: $t(112) = 2.64$, $p = .01$; Cohen's $d_z = .25$ (small)** **OCB: $t(112) = -3.30$, $p = .001$; Cohen's $d_z = -.31$ (small)**
Response Time	JS: $t(672) = .13$, $p = .899$ OC: $t(672) = .21$, $p = .838$ OCB: $t(672) = .40$, $p = .69$	**JS: $t(4) = -3.22$, $p = .032$; Cohen's $d_z = -1.44$ (very large)** OC: $t(4) = -2.42$, $p = .073$ **OCB: $t(4) = -4.18$, $p = .014$; Cohen's $d_z = -1.87$ (very large)**
Individual Reliability	JS: $t(448) = 1.29$, $p = .198$ OC: $t(448) = .91$, $p = .361$ OCB: $t(448) = -.09$, $p = .929$	JS: $t(228) = -1.67$, $p = .097$ OC: $t(228) = -1.35$, $p = .18$ OCB: $t(228) = .13$, $p = .899$
RIR	JS: $t(447) = 1.18$, $p = .24$ OC: $t(447) = 1.16$, $p = .245$ OCB: $t(447) = .90$, $p = .371$	JS: $t(229) = -1.63$, $p = .104$ OC: $t(229) = -1.83$, $p = .069$ OCB: $t(229) = -1.38$, $p = .17$
Instructed Response Items and Individual Reliability	JS: $t(365) = .95$, $p = .341$ OC: $t(365) = -.02$, $p = .986$ OCB: $t(365) = 1.24$, $p = .214$	JS: $t(128) = -1.04$, $p = .298$ OC: $t(128) = .02$, $p = .983$ OCB: $t(128) = -1.34$, $p = .181$
Instructed Response Items and RIR	JS: $t(377) = .86$, $p = .39$ OC: $t(377) = .31$, $p = .759$ **OCB: $t(377) = 2.11$, $p = .036$; Cohen's $d_z = .11$ (very small)**	JS: $t(299) = -1.01$, $p = .315$ OC: $t(229) = -.39$, $p = .697$ **OCB: $t(229) = -2.56$, $p = .011$; Cohen's $d_z = -.15$ (very small)**
Response Time and Individual Reliability	JS: $t(445) = 1.43$, $p = .154$ OC: $t(445) = 1.15$, $p = .251$ OCB: $t(445) = .25$, $p = .804$	JS: $t(231) = -1.83$, $p = .069$ OC: $t(231) = -1.64$, $p = .102$ OCB: $t(231) = -.33$, $p = .74$
Response Time and RIR	JS: $t(444) = 1.31$, $p = .191$ OC: $t(444) = 1.40$, $p = .163$ OCB: $t(444) = 1.25$, $p = .212$	JS: $t(232) = -1.80$, $p = .073$ **OC: $t(232) = -2.13$, $p = .035$; Cohen's $d_z = -.14$ (very small)** OCB: $t(232) = -1.79$, $p = .074$
Instructed Response Items and Response Time	JS: $t(562) = -.10$, $p = .925$ OC: $t(562) = -.78$, $p = .433$ OCB: $t(562) = 1.89$, $p = .059$	JS: $t(114) = .25$, $p = .802$ **OC: $t(114) = 2.06$, $p = .042$; Cohen's $d_z = .19$ (small)** **OCB: $t(114) = -3.63$, $p < .001$; Cohen's $d_z = -.34$ (small)**
Instructed Response Items, Response Time, and Individual Reliability	JS: $t(363) = 1.03$, $p = .303$ OC: $t(363) = .18$, $p = .86$ OCB: $t(363) = 1.62$, $p = .106$	JS: $t(313) = -1.12$, $p = .262$ OC: $t(313) = -.21$, $p = .837$ OCB: $t(313) = -1.65$, $p = .101$
Instructed Response Items, Response Time, and RIR	JS: $t(375) = .94$, $p = .35$ OC: $t(375) = .50$, $p = .617$ **OCB: $t(375) = 2.49$, $p = .013$; Cohen's $d_z = .13$ (very small)**	JS: $t(301) = -1.09$, $p = .277$ OC: $t(301) = -.62$, $p = .536$ **OCB: $t(301) = -2.85$, $p = .005$; Cohen's $d_z = -.16$ (very small)**

Note: Significant findings are in boldface.

	Regressions: Chow test (Careful/Careless)
Instructed Response Items	**JS-OCB: *F**= 11.022, p < .001** **OC-OCB: *F**= 13.701, p < .001**
Response Time	**JS-OCB: *F**= 9.77, p < .001** **OC-OCB: *F**= 8.226, p < .001**
Individual Reliability	**JS-OCB: *F**= 4.045, p = .018** OC-OCB: *F**= 1.727, p = .179
RIR	**JS-OCB: *F**= 3.333, p = .036** OC-OCB: *F**= 2.79, p = .062
Instructed Response Items and Individual Reliability	JS-OCB: *F**= 2.989, p = .051 **OC-OCB: *F**= 3.269, p = .039**
Instructed Response Items and RIR	**JS-OCB: *F**= 5.751, p = .003** **OC-OCB: *F**= 7.472, p = .001**
Response Time and Individual Reliability	JS-OCB: *F**= 2.765, p = .064 OC-OCB: *F**= .453, p = .636
Response Time and RIR	**JS-OCB: *F**= 3.187, p = .042** OC-OCB: *F**= 1.857, p = .157
Instructed Response Items and Response Time	**JS-OCB: *F**= 15.021, p < .001** **OC-OCB: *F**= 17.535, p < .001**
Instructed Response Items, Response Time, and Individual Reliability	**JS-OCB: *F**= 3.191, p = .042** **OC-OCB: *F**= 3.266, p = .039**
Instructed Response Items, Response Time, and RIR	**JS-OCB: *F**= 6.857, p = .001** **OC-OCB: *F**= 7.792, p <.001**

Note: Significant findings are in boldface.

	Regressions: Fit statistics	
	Full/Careful	Full/Careless
Instructed Response Items	JS-OCB: R^2=.10; adjusted R^2=.10; SEE=.50; β=.316** (minimal) OC-OCB: R^2=.17; adjusted R^2 =.17; SEE=.48; β=.417** (small)	JS-OCB: R^2=.19; adjusted R^2=.18; SEE=.51; β=.432** (small) OC-OCB: R^2=.13; adjusted R^2=.12; SEE=.53; β=.353** (small)
Response Time	JS-OCB: R^2=.11; adjusted R^2=.10; SEE=.50; β=.323** (minimal) OC-OCB: R^2=.14; adjusted R^2 = .14; SEE=.49; β=.377** (minimal)	**JS-OCB: R^2=.01; adjusted R^2= –.32; SEE=.68; β= –.10 (large)** **OC-OCB: R^2=.31; adjusted R^2=.08; SEE=.56; β=.558 (large)**
Individual Reliability	JS-OCB: R^2=.15; adjusted R^2=.15; SEE=.50; β=.389** (small) OC-OCB: R^2=.19; adjusted R^2 = .18; SEE=.49; β=.432** (small)	JS-OCB: R^2=.05; adjusted R^2=.04; SEE=.52; β=.216** (small) OC-OCB: R^2=.09; adjusted R^2=.08; SEE=.51; β=.295** (small)
RIR	JS-OCB: R^2=.14; adjusted R^2=.14; SEE=.51; β=.373** (small) OC-OCB: R^2=.19; adjusted R^2 = .19; SEE=.50; β=.436** (small)	JS-OCB: R^2=.05; adjusted R^2= .04; SEE=.49; β= .218** (small) OC-OCB: R^2=.07; adjusted R^2=.06; SEE=.49; β=.257** (small)
Instructed Response Items and Individual Reliability	JS-OCB: R^2=.15; adjusted R^2=.15; SEE=.49; β=.392** (small) OC-OCB: R^2=.21; adjusted R^2 = .21; SEE=.48; β=.455** (small)	JS-OCB: R^2=.06 adjusted R^2=.06; SEE=.53; β=.244** (small) OC-OCB: R^2=.09; adjusted R^2=.09; SEE=.52; β=.302** (small)
Instructed Response Items and RIR	JS-OCB: R^2=.13; adjusted R^2=.13; SEE=.52; β=.366** (small) OC-OCB: R^2=.20; adjusted R^2 = .20; SEE=.50; β=.448** (small)	JS-OCB: R^2=.07; adjusted R^2=.07; SEE=.50; β=.262** (small) OC-OCB: R^2=.09; adjusted R^2=.08; SEE=.49; β=.293** (small)
Response Time and Individual Reliability	JS-OCB: R^2=.15; adjusted R^2=.15; SEE=.49; β=.386** (small) OC-OCB: R^2=.17; adjusted R^2=.17; SEE=.49; β=.418** (small)	**JS-OCB: R^2=.05 adjusted R^2=.05; SEE=.54; β=.224** (moderate)** OC-OCB: R^2=.11; adjusted R^2=.10; SEE=.52; β=.327** (small)
Response Time and RIR	JS-OCB: R^2=.14; adjusted R^2=.14; SEE=.51; β=.369** (small) OC-OCB: R^2=.18; adjusted R^2=.18; SEE=.50; β=.422** (small)	**JS-OCB: R^2=.05; adjusted R^2= .05; SEE=.51; β= .227** (moderate)** OC-OCB: R^2=.09; adjusted R^2=.08; SEE=.49; β=.293** (small)
Instructed Response Items and Response Time	JS-OCB: R^2=.10; adjusted R^2=.10; SEE=.49; β=.314** (minimal) OC-OCB: R^2=.17; adjusted R^2 = .16; SEE=.474; β=.407** (small)	JS-OCB: R^2=.20 adjusted R^2=.19; SEE=.54; β=.443** (small) OC-OCB: R^2=.17; adjusted R^2=.17; SEE=.55; β=.417** (small)
Instructed Response Items, Response Time, and Individual Reliability	JS-OCB: R^2=.15; adjusted R^2=.15; SEE=.48; β=.39** (small) OC-OCB: R^2=.20; adjusted R^2=.19; SEE=.47; β=.441** (small)	JS-OCB: R^2=.06; adjusted R^2=.06; SEE=.54; β=.249** (small) OC-OCB: R^2=.11; adjusted R^2=.10; SEE=.53; β=.325** (small)
Instructed Response Items, Response Time, and RIR	JS-OCB: R^2=.13; adjusted R^2=.13; SEE=.50; β=.363** (small) OC-OCB: R^2=.19; adjusted R^2 = .19; SEE=.49; β=.435** (small)	JS-OCB: R^2=.07; adjusted R^2=.07; SEE=.51; β= .267** (small) OC-OCB: R^2=.10; adjusted R^2=.10; SEE=.50; β=.318** (small)

Note: Significant findings are in boldface. * $p < .05$. ** $p < .01$.

The Search for Mechanisms in Business Research: Reflections on Retroductive Analysis in a Multilevel Critical Realist Case Study

Deepak Saxena
Trinity College Dublin, Dublin, Ireland
saxenad@tcd.ie

Abstract: Many introductory research methods textbooks still divide the research paradigms into two broad approaches – positivist/quantitative/deductive and interpretive/qualitative/inductive. However, this bifurcation of research orientation does not do justice to the philosophical and methodological pluralism present within business research. This paper offers a third way by reflecting on a retroductive analysis in a critical realist case study. The philosophy of critical realism employs retroductive analysis to search for mechanisms underpinning the empirically observed events. Mechanism-based theorising is a suggested way in the business research to develop middle-range 'sometimes true' theories. This paper demonstrates the process of retroduction for the identification of mechanisms through an illustration of the data collection, coding and analysis process in a multilevel critical realist case study. In the process, it outlines the challenges faced and offers suggestions to overcome those challenges. This paper does not claim to provide a set of best practices for multilevel retroductive analysis. However, it is hoped that it sensitises business researchers to explore the critical realist perspective and to employ retroduction for mechanism-based theorising.

Keywords: Critical Realism, mechanism, retroduction, multilevel, coding, analysis

1. Introduction

Although there is a multiplicity of research philosophies in business research, many research methods textbooks still divide the philosophical discussion into two major approaches – positivism and interpretivism – which are usually presented as polar opposites (Knox, 2004). While the positivists primarily rely on deductive logic and test the hypothesis mostly using quantitative data, interpretivists usually look for meaning and patterns following inductive logic mostly using qualitative data. Consequently, business researchers working within the positivist philosophy claim to provide causally adequate theory via statistical analysis and those following interpretivist philosophy claim to provide adequate explanation at the level of meaning (Coldwell, 2007). Although there is some acknowledgement of other approaches in business research (e.g. design science or critical theory), the majority of the discussion still revolves around the two approaches. However, this bifurcation of research orientation does not do justice to the philosophical and methodological pluralism present within business research.

Critical realism is often proposed (Fleetwood, 2005; Mingers, 2004) as a third way in business research. This paper aims to sensitise business researchers towards exploring the critical realist perspective and employing retroduction for mechanism-based theorising. Towards this aim, the paper has twin objectives. First objective is to clarify the process of retroductive analysis by giving a systematic account of the data collection, coding and analysis process. Second objective is to discuss the challenges encountered during a retroductive analysis and offer suggestions to overcome those challenges. The paper is based on the observations and reflections during a critical realist case study within the enterprise system (ES) domain. Formative notes were prepared at different stages of the research and a summative reflection was conducted at the end of the research before writing this paper. The remainder of the paper is structured as follows. The following section provides a brief summary of the critical realist philosophy and retroductive analysis. Section 3 outlines the key requirements of data collection for a critical realist case study and reflects on how those requirements were fulfilled. Section 4 discusses the identification of mechanisms using multilevel retroductive analysis. It also outlines the challenges in retroductive analysis and offers suggestions to overcome those challenges. Finally, section 5 concludes the paper by discussing key takeaways and future research.

2. Critical realist philosophy and retroductive analysis

Before the exposition of critical realist philosophy and its methodological considerations, perhaps it is pertinent to clarify few assumptions here. First, it is acknowledged here that there is not a neat and tidy distinction between various research paradigms and the choices a researcher makes usually lie on a continuum (Cunliffe, 2011; Walsh et al., 2015a). Second, it is not asserted here that critical realism is a panacea for all

Reference this paper: Saxena, D., 2019. The Search for Mechanisms in Business Research: Reflections on Retroductive Analysis in a Multilevel Critical Realist Case Study. *The Electronic Journal of Business Research Methods,* 17(1), pp. 17-27, available online at www.ejbrm.com

business research or that it is free from any defects. Third, it is acknowledged that there is no 'single' version of critical realism and that it has some linkages with the pragmatist approach (DeForge and Shaw, 2012; Johnson and Duberley, 2000; Lipscomb, 2011). This research acknowledges the existence of a variety of different epistemological positions (within and outside critical realism) that have their own distinctive ways of engaging with reality and undertaking research (Johnson and Duberley, 2000; Newton, Deetz and Reed, 2011). Consequently, the aim for the researchers should be to maintain consistency with regard to the epistemological assumption they deploy. The main motive of this section is to introduce the reader with the key assumptions of the critical realist position underpinning this study. It may be noted, however, that this is not a place to discuss the critical realist philosophy in its entirety and an interested reader is referred to the works cited in this paper.

2.1 Basic tenets of critical realism

Roy Bhaskar is often credited with the development of critical realist philosophy initially for natural sciences (Bhaskar, 1975) and then for social sciences (Bhaskar, 1989). The critical realist philosophy is *realist* in the sense that it assumes that social and natural reality exist independently of our cognitive processes – an extra-mental reality exists whether or not human beings can actually gain cognitive access to it (Fleetwood, 2005; Johnson and Duberley, 2000). This may be a concrete reality (e.g. atmospheric pressure), or an abstract reality (e.g. social pressure), but it exists irrespective of our understanding of it (Maxwell, 2012). However, unlike the naive realist ontology of positivism that reality exists and is completely apprehendable, critical realists acknowledge that it is imperfectly apprehendable due to our bounded rationality (Simon, 1982) and fundamentally intractable nature of phenomena (Burrell and Morgan, 1979). The philosophy is considered *critical* since it asserts that the claims about reality are to be accepted with the caveat that there are significant limitations of objectivity of our knowledge (Mingers, Mutch and Willcocks, 2013). This position defends critical realism both against logical positivism that would reduce the world to empirically observed measurable events, and various forms of postmodernisms that would reduce the world to our ideas (Ackroyd and Fleetwood, 2000; Fleetwood, 2004, 2005; Maxwell, 2012; Mingers et al., 2013; Reed, 2005). Following Cunliffe's (2011) typology, a critical realist study would fall within the objectivist knowledge problematic but does not fall prey to epistemic fallacy (Mingers, 2004) of conflating reality with our understanding of it.

As originally proposed by Bhaskar (Bhaskar, 1975, 1989; Collier, 1994; Sayer, 2000, 2010), critical realism stratifies reality into three nested domains – real, actual, and empirical. The domain of the *real* is conceived to be independently existing and includes the entities, the structures and the causal powers inherent to them. The domain of the *actual* is considered a subset of the real and includes the events that occur due to the enactment of the causal powers of structures and entities. These events may or may not be observed by humans. Finally, the domain of the *empirical* is understood to be a subset of the actual and consists of our experiences via measurements or sense perception. The purpose of a critical realist study is to explain a given set of events by uncovering underlying mechanisms which, if they existed and were enacted, could have produced these events (Bhaskar 1975, 1989). For this reason, mixed-method research is often recommended for critical realist studies. While quantitative data helps in determining empirical regularities (Tsang, 2014), qualitative data helps in identifying the mechanisms that emerge from the components of a physical and social structure to produce the events of interest (Sayer, 2000, 2010). However, critical realist perspective also recognises that each event is not only dependent on the causal powers available within a social structure, but also on the continuously changing contextual conditions and the evolving properties of the structure (Wynn and Williams, 2012). Therefore, a causal explanation in critical realism accounts for a set of existing and enacted mechanisms, along with the impact of any structural factors and contextual conditions that generated the outcome being studied (Wynn and Williams, 2012).

2.2 Realism in business research

Within the domain of business research, although there was some discussion on realism at times (e.g. Dubin, 1982), serious attention was given towards the turn of the century. Tsang and Kwan (1999) discuss the potential of critical realism for replication and theory development in business research. Claiming a 'realist turn' in organisational and management studies, Ackroyd and Fleetwood (2000) and Fleetwood and Ackroyd (2004) present various theoretical, methodological, and empirical works associated with realism, arguing for explicit application of the realist paradigm. It is argued (Reed, 2005; Wynn and Williams, 2012) that due to its focus on an extra-mental reality, critical realism offers a way to address the rigor–relevance gap in business research. It is also interesting to note that independent of the discussion on critical realism, there are parallel developments in the idea of social mechanisms (McGrath, 2013). Although Coleman (1964) introduced the concept of social mechanism as 'sometimes-true theory', it was explored in more detail in Hedström and Swedberg (1998). Davis (2006) and Avgerou (2013) suggest that the notion of mechanisms is implicit in many theories, for example, isomorphic 'forces' in the institutional theories. Campbell (2005) discusses various environmental, cognitive, and relational mechanisms to explain the process of institutional change. Mechanism are seen as building-blocks of middle-range theories that are useful for explaining problems around organisations and organising and to form more general process theories (Pettigrew, Woodman and Cameron, 2001). Social mechanisms are about "the wheelwork or agency by which an effect is produced. In this way, mechanisms do not merely address what happened but also how it happened" (Hernes, 1998, p. 74). Anderson et al. (2006) argue that that a focus on mechanisms enables one to move beyond individual variables and their linkages to consider the bigger picture of action in its entirety. A mechanism-based explanation moves beyond *describing what* to *explaining how*, and thereby clarifies causal ambiguity (Pajunen, 2008) surrounding a phenomena. In this sense, the philosophy of critical realism provides a sound basis for mechanism-based theorising within business research. In the ES domain, the mechanism-based theorising is considered even more relevant because the domain is marred by identification of too many surface-level critical success factors (Saxena and McDonagh, 2017) with less than adequate understanding of individual factors (Martin and Huq, 2007) under consideration.

2.3 Retroductive analysis in critical realism

Researchers usually engage in retroductive analysis (Chiasson, 2005; Danermark et al., 2002; Tsang, 2014) iteratively during a critical realist study. The use of the term requires some clarification here (thanks to the reviewer for pointing this aspect). Charles Pierce uses the term 'abduction' and 'retroduction' interchangeably in his earlier writings, but seems to differentiate between the two in his later work (Chiasson, 2005). In the latter interpretation, abduction is considered a mode of inference on par with induction and deduction. While Chiasson (2005) interprets abduction as making a 'hunch' on underlying mechanism, for Danermark et al. (2002) the key element of abduction is redescription and recontextualisation of structure and events. Conceiving abduction in this way, retroduction is defined as a recursive application (Chiasson, 2005) of abduction, induction, and deduction. Retroduction is more iterative and creative in nature as the researcher moves back and forth between the data and explanation. As Danermark et al. (2002) suggest, it involves *transfactual* thinking because we need to think beyond the observed factual events and theorise on underlying mechanisms. Retroduction is about advancing from the empirical observation of events to a conceptualisation of mechanism and contextual conditions that produce those events (Johnson and Duberley, 2000). Through an iterative process, researcher improves the understanding of those mechanisms (Tsang, 2014; Volkof and Strong, 2013).

Since retroduction is largely a creative and intuitive process (Danermark et al., 2002), issuing specific guidelines for retroducing mechanisms is considered problematic (Wynn and Williams, 2012). However, Tsang (2014) notes four types of retroduction based on Eco (1983) – overcoded, undercoded, creative, and meta-retroduction – that may help in identifying the mechanisms. In *overcoded* retroduction, the mechanisms are directly available from the literature and the researcher's task is to explain the events employing those mechanisms. In *undercoded* retroduction, the current body of knowledge suggests a number of potential mechanisms and the researcher determines the ones that best explain the events under consideration. In *creative* retroduction, the researcher has to invent the mechanism because no suitable mechanisms are available in the literature. Finally, in *meta-retroduction*, observations do not fit our current conceptual schema and require us to think anew. This may result in some kind of paradigm change (Kuhn, 1962). The retroduction process for this research is discussed in detail in section four. Before that, however, next section discusses the data collection and preparation for a critical realist case study.

3. Data collection and preparation for a critical realist case study

One implication of critical realist data collection is that the collected data needs to be longitudinal to enable search for patterns across processes (Pettigrew, 1997; Dawson, 1997; Langley et al., 2013). George and Bennett (2005: p. 21) note that if performed with sufficient details, process tracing through a single longitudinal case study helps in examining the operation of causal mechanisms. Single case design allows the researcher to look for many intervening factors and inductively observe any unexpected aspects of the operation of a particular causal mechanism, or to identify the contextual conditions that activate the causal mechanism. Therefore, the main challenge for this study was to look for and gain access to a case that allows a longitudinal study of the ES implementation. After exploring multiple organisations, the researcher gained access to a public service organisation from the health sector. The case organisation engaged in three instances of ES implementations during 2000-2015, with the first and the third implementation deemed successful by the organisation. To begin with, the secondary data associated with the case was collected and analysed. This included both the documentation made publicly available by the case organisation as well as the archival data from other public sources. Analysis of the documents served three purposes. First, it enabled the researcher to arrive at a chronology of events. Second, it guided towards the identification of key actors who would be interview participants. Finally, it sensitised the researcher towards certain themes around which the interview questions were framed. Overall, the secondary data analysis provided a scaffolding to the primary data collection. Primary data mainly consisted of in-depth qualitative interviews with organisational participants. This study adopted responsive interviewing method of Rubin and Rubin (2012) in which interview starts with general questions and new questions are framed based on the participants' answers. This allows capturing rich details as they emerge. Additionally, further secondary data collection continued along with the interviews.

Another implication of critical realist data collection is that data collection needs to be multilevel to search for mechanisms operating across multiple levels (Langley, 1999; Pettigrew, Woodman and Cameron, 2001), for example, for assessing the impact of a changing socio-economic context on features of intra-organisational processes. Since this research primarily focused on technology in organisations, the three levels suggested by the socio-technical systems theory – macrosocial, organisation, and work-system level (Trist, 1981; Winter, Berente, Howison and Butler, 2014) – were employed for data collection. *Work-systems* carry out a specific set of activities in an identifiable and bounded subsystem of an organisation such as a department or a service unit. It may consist of a group of personnel and the relevant equipment and other resources. *Organisational* systems may correspond to a plant, a workplace, or a public agency. Finally, *macrosocial* systems include communities, industrial sectors, and institutions operating at the overall level of a society.

A main challenge at this stage was to ensure the multilevel nature of data. Relevant documents pertaining to all three levels were collected. At the work-system level, relevant documents included system requirement specifications, project planning documents, minutes of the project meetings, and project audit documents. Organisational level documentation included strategy documents, minutes of the board meetings, and annual reports. Finally, at the macrosocial levels, there were reports of the government auditor, minutes of parliamentary committee meetings, and newspaper articles. Similarly, during the primary data collection, interview participants included the members from the project team (both IT and non-IT), middle level management, and the top-level management to ensure the internal generalisability (Maxwell, 2012). The top management was considered a proxy for assessing the macrosocial processes since they often interact with organisation's external context (Martin and Huq, 2007). Ideally, the actors outside the organisation (e.g. people from the Department of Health, the ES supplier) should have been interviewed for comprehensively accessing the macrosocial level, it was not possible due to access and time limitations. To overcome this limitation, secondary data also included publicly available information on the Department of health and the ES supplier.

Once all the interviews (twenty-seven in total) were conducted and transcribed, the next challenge was to make sense of everything. First, a descriptive case narrative was written describing the events, structure and context. The narrative was in the form of process tracing which was highly specific to the case (George and Bennett, 2005, p. 210) and did not make any explicit use of the theory. Essentially, it focused on providing 'pure' description (Walcott, 2009, p. 27-29) trying to present the events as they unfolded without any efforts to analyse so that richness of the case is not compromised. It was a major challenge to free the narrative from any attempt of theorising. For this reason, writing of the narrative was commenced immediately after the data collection, i.e. in parallel with the transcription of the interviews. The narrative was continuously revised based on the emergent understanding of the researcher. Keeping the case narrative separate from the analysis also helped in ensuring the descriptive validity (Maxwell, 2012) of the narrative. The strategy of temporal bracketing (Langley, 1999) was used in writing the case narrative. For this purpose, key events at all three levels (macrosocial, organisational, and work-system) were identified across the timeline. Subsequently, the event-sequence and the interconnections between the events were narrated in detail based on the primary and secondary data. Following the socio-technical perspective, there was an effort to focus on both the social and the technical aspects of the exercise. This helped in presenting a holistic picture of the ES implementations in the case organisation. Once the writing of the narrative was complete (i.e. all events were described in sufficient detail), the next step was the identification of mechanisms using retroductive analysis, which is explained in the next section.

4. Identification of mechanisms using retroductive analysis

A three-stage process (Farquhar, 2012; Miles, Huberman and Saldaña, 2013; Saldaña, 2013) of coding and analysis was followed for the identification of mechanisms. Since the coding process is similar to the one recommended in the grounded theory, it needs some clarification here. Although grounded theory is sometimes seen as a strictly inductive approach for analysing qualitative data (Douglas, 2003), over the years it is evolved as a more general and flexible method of data analysis within diverse research paradigms (Hunter et al., 2005; Walsh et al., 2015b). However, the analysis process is not claimed here to be classic grounded theory (Glaser and Strauss, 1967) since it was not an explicit part of the research design and because there was a continuous to and fro movement between data and theory. However, the process described in this section might suitably be termed as *grounded inquiry* (Fendt and Sachs, 2008) or *grounded theorising* (Holton in Walsh et al, 2015b). The analysis and coding process unfolded in three stages using a dialectic process of retroductive reasoning involving method and creativity (Klag and Langley, 2013).

4.1 Using induction to generate first-order codes

The coding process started with the first-order coding in which the descriptive codes were directly assigned to the chunk of data. The orthodox reading of the grounded theory method recommends microanalysis using line-by-line or even word-by-word coding. However, it may result in too much confusion for the researcher (Allen, 2003). Therefore, following Miles, Huberman and Saldaña (2013), the selected chunk of data was of varying lengths depending on the richness of the data and its context. All interview transcripts and relevant secondary data were coded at this stage. There were twin challenges at this stage. Being an *inductive* exercise, the codes had to be based on the data and not on the theory. A second and more crucial challenge was to preserve the multilevel nature of the research. In first-order coding, descriptive coding was applied along with nested coding (Saldaña, 2013). Descriptive coding summarised the chunk of text in a word or short phrase. Here, care was taken to remain true to the data and not to force any specific theoretical terminology. Once a first-order code was assigned to a block of text, the nested coding was used to preserve the contextual and multilevel nature of the data. Since the study used the socio-technical systems theory as a sensitising device, all the descriptive codes were assigned to one of the levels (Trist, 1981; Winter et al., 2014) – macrosocial, organisational, or the work-system level. The level of the code was determined based on the entities/events being discussed in the text and how they aligned with the definitions provided by Trist (1981). Depending on the research problem, business researchers may use other levels such as micro-meso-macro (Papadimitriou, 2010) or individual-group-organisational (Rice, 1969) etc. Table 1 illustrate a couple of examples from each level to demonstrate the first-order coding process.

Table 1: Example first-order codes

Text Block	Code	Level
"...they would appear to have about 70 percent of the blood banking market worldwide, so, absolutely a dominant player."	Dominant supplier	Macrosocial
"So when [enterprise system] came-in and we ended up before Public Accounts Committee trying to explain the overrun in cost."	Public accountability	Macrosocial
"Our concern in scientific side is safety. Have we tested it and is everything right? And is the donation fit to go?"	Safety Culture	Organisational
"I think [case organisation] has to understand that IT is an enabling service. Its role is to enable everybody to do their job, not controlling them..."	Role of IT	Organisational
"There was an unmerciful push to say go-live at all costs and that was driven completely by [project sponsor] and [project manager]."	Focus on go-live	Work-system
"So the version that we are taking to go-live with is a lot more stable, without a doubt."	System maturity	Work-system

4.2 Using deduction to generate second-order codes

Once all data was coded in the form of first-order codes, the first-order codes were grouped into second-order codes based on their underlying similarity. This phase introduced the *deductive* component of the analysis since the majority of second order codes were drawn from existing literature on ES critical success factors (Finney and Corbett, 2007; Saxena and McDonagh, 2017). During this exercise, the corresponding level of the first-order codes was maintained, thereby assigning levels to second-order codes as well. Table 2 illustrates the second-order coding process using an example from each level. It is acknowledged here that procedure was not one-way and these codes were subject to revision, merger or deletion based on researcher's emerging understanding of the constructs (Miles, Huberman and Saldaña, 2013).

Table 2: Example second-order codes

First-order codes/Level	Second-order code	Level
Being First User/ Macrosocial	ES Market Structure	Macrosocial
Dominant Supplier/ Macrosocial		
Niche Market/ Macrosocial		
System Usage in Industry/ Macrosocial		
Business Integration/ Organisational	Business Vision	Organisational
Fragmented view of the Organisation/ Organisational		
IS Strategy/ Organisational		
Quality Focus/ Organisational		
Risk Aversion in Case Organisation/ Organisational		
Safety Focus/ Organisational		
Operational Improvement/ Organisational		
Operational Inefficiency/ Organisational		
Project Initiation/ Organisational		
Blood Control System/ Work-system	ES Artefact	Work-system
Data Migration/ Work-system		
Database Integration/ Work-system		
e-Financials/ Work-system		
Implementing BOSS/ Work-system		
Interim Label / Work-system		
Introducing ISBT-128/ Work-system		
MIS Reporting/ Work-system		
Special Testing Barcode/ Work-system		
System Complexity/ Work-system		
System Configuration/ Work-system		
System Constraints/ Work-system		
System Extension/ Work-system		
System Functionality/ Work-system		
System Maintenance/ Work-system		
System Maturity/ Work-system		
System Performance/ Work-system		
System Quality/ Work-system		
System Security/ Work-system		
System Terminology / Work-system		

4.3 Retroducing the mechanisms and explanation building

The final stage focused on the identification of underlying mechanisms and theory development. Compared to the coding process, this stage was more iterative and creative in nature as the analysis moved back and forth between the data and explanation. Along with the identification of mechanisms, explanation building goes hand-in-hand in a critical realist analysis. This is what George and Bennett (2005) call 'analytical process tracing'. This involved converting the codes and the descriptive narrative into an analytical causal explanation presented in explicit theoretical form. The goal here was to build an explanation of the events on *how* or *why* something happened. This phase involved a series of iterations and revisions to achieve greater explanation and theoretical coherence (Yin, 2013: p. 149). The objective was to identify the most complete and logically compelling explanation of the observed events given the specific conditions of the contextual environment (Wynn and Williams, 2012). A significant challenge at this stage was to ensure theoretical validity and (theoretical) generalisation when developing and explaining the theoretical framework. This meant ensuring that the identified mechanisms were logically consistent, were based on the empirical evidence, and were generally supported by the literature (Maxwell, 2012). However, it may be noted here that in some cases, the researcher might need to use creative or meta-retroduction (Eco, 1983) because existing theory does not offer any suitable mechanism.

Explanation building initially started with overcoded retroduction (Eco, 1983; Tsang, 2014) based on the constructs from the literature review. During the literature review it was conceived (*abduction* as in Chiasson, 2005) that the four process-mechanisms of Van de Ven and Poole (1995) – lifecyle, teleology, dialectic, and evolution – would neatly fit the bill. However, many events remained unexplained when using the four mechanisms. As an alternative, Anthony Giddens's (1984) structuration theory was explored since there was some evidence of legitimation process. However, again the explanation remained lacking due to a relative lack of data at the macrosocial level. Moreover, there was not enough evidence for other two mechanisms – signification and domination – noted in the structuration theory. Therefore, a major challenge at this stage was to remain true to the emergent understanding and not to force any predetermined theory. This motivated the researcher to look for mechanisms across diverse literature-sets, including organisational theory and industrial economics, thereby using undercoded retroduction (Eco, 1983; Tsang, 2014). In practice, this involved looking for commonality among the second-order codes within and across levels, going back to the literature in search of suitable mechanisms that underpin the commonality, and redescription and recontextualisation of the events under consideration (*abduction* as in Danermark et al., 2002). Many candidate mechanisms were considered before finally arriving at four underlying mechanisms that formed the building blocks of the final theoretical framework. Table 3 shows the finally arrived key mechanisms and associated second-order codes with corresponding levels.

Table 3: Identification of underlying mechanism

Second-order codes/Level	Mechanism	Level
Enterprise System Market Structure/ Macrosocial	Market Mechanism	Macrosocial
IS Resource Market/ Macrosocial		
Supplier's Push/ Macrosocial		
Support from Supplier/ Macrosocial		
Business Visions/ Organisational	Institutionalisation	Organisational
Change Management/ Organisational		
Organisational Learning/ Organisational		
Other Organisational Exercises/ Organisational		
Project Leadership/ Organisational		
Project Ownership/ Organisational		
Role of IT/ Organisational		
Superusers Selection/ Organisational		
User Engagement/ Organisational		
User Exposure/ Organisational		
Workarounds/ Organisational		
BPR-Customisation/ Work-system	Affordance Mechanism	Work-system
Enterprise System Artefact/ Work-system		
Gap Analysis/ Work-system		
Hardware-Interfacing / Work-system		
IS Infrastructure/ Work-system		
Partner's IS Infrastructure/ Work-system		
Superusers' Skills/ Work-system		
External Control/Macrosocial	Control Mechanism	Multilevel
Institutional Context/Macrosocial		
Change Control/ Organisational		
Internal controls/ Organisational		
Top Leadership/ Organisational		
Change Control/ Work-system		
Project Controls/ Work-system		

It is to be noted here that while all four mechanisms are based in literature, existing literature may not use the terms 'mechanism' when using similar concepts (Avgerou, 2013; Davis, 2006), perhaps avoiding an explicit critical realist stance. Of the four mechanisms that are identified in this study, only affordance mechanism is distinctly identified as a critical realist mechanism (Volkoff and Strong, 2013) as such. *Affordance mechanism* captures as action possibilities and opportunities that emerge from actors engaging with the system (Faraj and Azad, 2012). An affordance perspective recognises how an object supports a set of business processes and constraints some other business processes (Zamutto, et al., 2007) when an organisation attempts to appropriate the technology. Since it includes aspects related to the technology and the business process, it primarily operates at the work-system level. The *institutionalisation mechanism* is conceived here as a process by which a social structure attains a stable and durable state or property (Currie, 2009) that produces recognisable, repetitive patterns of interdependent actions, carried out by multiple actors within a pre-existing social context (Feldman and Pentland, 2003). The *market mechanism* refers to product availability, market structure, buyer-supplier relationship, and the impact of these on the enterprise system lifecycle. The market mechanism was a dominant mechanism at the macrosocial level. Finally, the *control mechanism* is understood as a set of activities that are conducted in a project to regulate or adjust the behaviour of the stakeholders, to motivate participants, and to ensure that their capabilities are fully applied to advance the ES initiative towards its objectives (Kirsch, 1997, 2004). The control mechanism operates at all three levels – work-system, organisational, and macrosocial level.

Once the underlying mechanisms were identified with a level of reasonable certainty, the last phase of the case study involved the development of an explanatory framework involving these mechanisms. In other words, the case narrative was explained in terms of the four mechanisms. The explanatory framework is not included here since it falls outside the scope of this paper.

4.4 Looking back – challenges and suggestions for applying retroduction

Based on earlier discussion, Table 4 summarises the steps, associated challenges, and offers suggestion for conducting a critical realist case study and applying retroduction for the identification of mechanisms. It is not claimed here to offer a set of best practice for applying retroduction for mechanism-based theorising. At best, these suggestions should be treated as pointers for further exploration by researcher employing the critical realist thinking.

Table 4: Challenges and Suggestion for Applying Retroduction

Steps	Challenges	Suggestions
Case selection and data collection	To include a longitudinal element in the case	Focus on the process tracing instead of focusing on a one-off event.
	To ensure multilevel data collection	Explicitly include the concept of levels in the research framework. Collect and sort data according to the pre-determined levels. Try to interview people across the hierarchy. If possible, talk to the people from outside the organisation. Collect secondary data on relevant entities/events even if they are outside the case organisation.
Presenting a 'pure' narrative	To free the narrative from any attempt of theorising	Start writing the narrative immediately after the data collection in parallel with the transcription of the interviews
Retroductive analysis	To retain the multilevel nature of the data	Use nested coding by assigning a dominant level to each code. Maintain the level of the code when grouping first-order codes into second-order codes, and second-order codes into mechanisms.
	Recursively apply abduction, induction and deduction	Make a 'hunch' on candidate mechanism based on the literature review (abduction). Assign first-order codes based on the data (induction). See if the first-order codes can be grouped into second-order codes corresponding to literature (deduction). Engage in recontextualisation and redescription using candidate mechanisms (abduction).
Identification of mechanisms and explanation building	Ensuring theoretical validity and (theoretical) generalisation	Do not try to impose specific mechanism/theory if your data does not support it. Be open to the possibility of using different mechanisms available in your and other disciplines. Engage in a series of iterations and revisions to achieve greater explanation and theoretical coherence.

5. Conclusion

There are calls in business research literature to look beyond the positivist and interpretivist traditions and to consider other research philosophies. Taking a cue from the realist turn (Ackroyd and Fleetwood, 2000; Fleetwood and Ackroyd, 2004) and the calls for mechanism-based theorising (Anderson et al., 2006; Avgerou, 2013; Campbell, 2005; Davis, 2006) in business research, this paper outlines the application of multilevel retroduction in a critical realist case study. The process of retroduction is demonstrated through an illustration of a multilevel coding and analysis process for the identification of mechanisms. A key implication of mechanism-based theorising is that it directs over focus away from the surface-level events and shifts it to underlying mechanisms that cause those events. The managerial implication is that managers can focus on interacting with underlying mechanisms rather than trying to manage empirical level events. Apart from ES domain, another ripe area for the mechanism-based theorising would be organisational change, paving a way for contextual and processual understanding of change (Pettigrew, 1997; Pettigrew, Woodman and Cameron, 2001).

The main takeaway from the paper is that the design and retroductive analysis in a critical realist case study should be sensitive to the longitudinal and multilevel elements of the research problem. During case selection, the focus should be on the process tracing instead of focussing on one-off events. It certainly helps during the data collection and analysis process that the research framework explicitly incorporates the concept of levels, such as the one employed in this study. The coding process should retain the predetermined levels of the first-order, second-order, and when identifying a mechanism, which could then be a single level or a multilevel mechanism. Since the process of retroduction includes employing abduction, induction, and deduction in a recursive manner, the process of developing explanation is iterative in a critical realist case study. Another learning is that the whole exercise should not be bound by disciplinary boundaries and should be open for revision based on emerging understanding of the researcher.

While this paper does not claim to provide a set of best practice for multilevel retroductive analysis, it hopes to sensitise (and possibly lure) business researchers to explore the critical realist perspective and to employ retroduction for mechanism-based theorising. Due to a technological element, this research used the socio-technical framework for data collection and analysis. Future research in this area could focus on the application of other multilevel frameworks depending on varied context. Furthermore, this paper did not explore the interaction between the mechanisms. Future research direction would focus on providing methodological guidelines and managing the research complexity when the interaction among mechanisms is being studied for a more comprehensive understanding of the phenomena.

References

Ackroyd, S. and Fleetwood, S. eds., 2000. *Realist perspectives on management and organisations*. London: Routledge.

Allan, G., 2003. A critique of using grounded theory as a research method. *Electronic journal of business research methods*, 2(1), pp.1-10.

Anderson, P.J., Blatt, R., Christianson, M.K., Grant, A.M., Marquis, C., Neuman, E.J., Sonenshein, S. and Sutcliffe, K.M., 2006. Understanding mechanisms in organizational research: Reflections from a collective journey. *Journal of Management Inquiry*, *15*(2), pp.102-113.

Avgerou, C., 2013. Social Mechanisms for Causal Explanation in Social Theory Based IS Research. *Journal of the Association for Information Systems* 14(8), pp.399-419.

Bhaskar, R., 1975. *A realist theory of science*. Leeds: Leeds Books Ltd.

Bhaskar, R., 1989. *The possibility of naturalism: a philosophical critique of the contemporary human sciences.* Hertfordshire: Harvester Wheatsheaf.

Burrell, G. and Morgan, G., 1979. *Sociological paradigms and organisational analysis* (Vol. 248). London: Heinemann.

Campbell, J. L., 2005. Where do we stand? Common mechanisms in organizations and social movements research. In: Davis, G.F., McAdam, D., Scott, W.R. and Zald, M.N. eds., 2005. *Social movements and organization theory*. Cambridge: Cambridge University Press. Ch.2.

Chiasson, P., 2005. Abduction as an aspect of retroduction. *Semiotica*, 2005(153-1/4), pp.223-242.

Coldwell D.A.L., 2007. Is Research that is Both Causally Adequate and Adequate on the Level of Meaning Possible or Necessary in Business Research? A Critical Analysis of some Methodological Alternatives. *Electronic Journal of Business Research Methods*. 5(1), pp.1 – 10.

Coleman, J. S., 1964. *Introduction to mathematical sociology*. New York: Free Press.

Collier, A., 1994. *Critical realism: an introduction to Roy Bhaskar's philosophy*. London: Verso.

Cunliffe, A. L., 2011. Crafting qualitative research: Morgan and Smircich 30 years on. *Organizational Research Methods*, 14(4), pp.647-673.

Currie, W., 2009. Contextualising the IT artefact: towards a wider research agenda for IS using institutional theory. *Information Technology & People*, 22(1), pp.63-77.

Danermark, B., Ekstrom, M., Jakobsen, L. and Karlsson, J. Ch,. 2002. *Explaining society: An introduction to critical realism in the social sciences*. London: Routledge.

Davis, G. F., 2006. Mechanisms and the theory of organizations. *Journal of Management inquiry*, *15*(2), pp. 114-118.

Dawson, P., 1997. In at the deep end: conducting processual research on organisational change. *Scandinavian journal of management*, 13(4), pp.389-405.

DeForge, R. and Shaw, J., 2012. Back-and fore-grounding ontology: exploring the linkages between critical realism, pragmatism, and methodologies in health & rehabilitation sciences. *Nursing Inquiry*, 19(1), pp.83-95.

Douglas, D., 2003. Inductive theory generation: A grounded approach to business inquiry. *The Electronic journal of business research methods*, 2(1), pp. 47-54.

Dubin, R., 1982. Management: Meanings, methods, and moxie. *Academy of Management Review*, 7, pp.372-379.

Eco, U., 1983. Horns, hooves, insteps. In: Eco, U., Sebeok, T.A. Eds. 1983, *The Sign of Three: Dupin, Holmes, Peirce*. Bloomington: Indiana University Press.

Faraj, S. and Azad, B., 2012. The materiality of technology: An affordance perspective. In Leonardi, P.M., Nardi, B.A. and Kallinikos, J. Eds., 2012. *Materiality and organizing: Social interaction in a technological world*. Cambridge: Cambridge University Press. pp.237-258.

Farquhar, J. D., 2012. *Case Study Research for Business*. London: Sage.

Feldman, M. S. and Pentland, B. T., 2003. Reconceptualizing organizational routines as a source of flexibility and change. *Administrative science quarterly*, *48*(1), pp.94-118.

Fendt, J. and Sachs, W., 2008. Grounded theory method in management research: Users' perspectives. *Organizational Research Methods*, *11*(3), pp.430-455.

Finney, S. and Corbett, M., 2007. ERP implementation: a compilation and analysis of critical success factors. *Business Process Management Journal*, 13(3), pp.329-347.

Fleetwood, S., 2004. The ontology of organisation and management studies. In: Ackroyd, S. and Fleetwood, S. Eds. 2004. *Realist Applications in Organisation and Management Studies*, London: Routledge.

Fleetwood, S., 2005. Ontology in organization and management studies: A critical realist perspective. *Organization*, *12*(2), pp.197-222.

Fleetwood, S. and Ackroyd, S. Eds. 2004. *Critical realist applications in organisation and management studies* (Vol. 11). London: Routledge.

George, A. L. and Bennett, A. 2005. *Case studies and theory development in the social sciences*. Cambridge: MIT Press, Massachusetts.

Glaser, B. and Strauss, A., 1967. *The discovery of grounded theory*, Chicago: Adeline, USA.

Giddens, A., 1984. *The constitution of society: Outline of the theory of structure*. Berkeley: University of California Press.

Hedström, P. and Swedberg, R., Eds. 1998. *Social mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press.

Hernes, G., 1998. Real virtuality. In Hedström, P. and Swedberg, R. Eds. 1998. Social mechanisms: An analytical approach to social theory. Cambridge: Cambridge University Press. pp.74-101.

Hunter, K., Hari, S., Egbu, C. and Kelly, J., 2005. Grounded theory: Its diversification and application through two examples from research studies on knowledge and value management. *The Electronic Journal of Business Research Methods*, 3(1), pp.57-68.

Johnson, P. and Duberley, J., 2000. *Understanding management research: An introduction to epistemology*. London: Sage.

Kirsch, L. J., 1997. Portfolios of control modes and IS project management. *Information systems research*, *8*(3), pp.215-239.

Kirsch, L. J., 2004. Deploying common systems globally: The dynamics of control. *Information systems research*, *15*(4), 374-395.

Klag, M. and Langley, A., 2013. Approaching the conceptual leap in qualitative research. *International Journal of Management Reviews*, *15*(2), pp.149-166.

Knox, K. T., 2004. A researcher's dilemma-philosphical and methodological pluralism. *The Electronic Journal of Business Research Methods*, 2(2), pp.119-128.

Kuhn, T.S., 1962. *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Langley, A. 1999. Strategies for theorizing from process data. *Academy of Management review*, *24*(4), pp.691-710.

Langley, A., Smallman, C., Tsoukas, H. and Van de Ven, A. H., 2013. Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of management journal*, 56(1), pp.1-13.

Lipscomb, M. (2011) Critical realism and realist pragmatism in mixed methods: Problematics of event identity and abductive inference. In: *American Educational Research Association Annual Meeting*, 2011, New Orleans, Louisiana, USA, 8th - 12th April, 2011. Available from: http://eprints.uwe.ac.uk/14188

Martin, T.N. and Huq, Z., 2007. Realigning top management's strategic change actions for ERP implementation: how specializing on just cultural and environmental contextual factors could improve success. *Journal of Change Management*, 7(2), pp.121-142.

Maxwell, J. A., 2012. *A realist approach for qualitative research*. London: Sage.

McGrath, K., 2013. The potential of generative mechanisms for IS research. In: Association for Information Systems, *Thirty Fourth International Conference on Information Systems*, Milan 2013.

Miles, M. B., Huberman, A. M. and Saldaña, J., (2013). *Qualitative Data Analysis*. London: Sage.

Mingers, J., 2004. Real-izing information systems: critical realism as an underpinning philosophy for information systems. *Information and Organization*, 14(2), pp.87-103.

Mingers, J., Mutch, A. and Willcocks, L., 2013. Critical realism in information systems research. *MIS Quarterly*, 37(3), 795-802.

Newton, T., Deetz, S. and Reed, M., 2011. Responses to social constructionism and critical realism in organization studies. *Organization studies*, *32*(1), pp.7-26.

Pajunen, K., 2008. The nature of organizational mechanisms. *Organization Studies*, 29(11), pp.1449-1468.

Papadimitriou, A., 2010. Looking for Clues about Quality: A Multilevel Mixed Design on Quality Management in Greek Universities. *The Electronic Journal of Business Research Methods*, 8(2), 85.

Pettigrew, A. M., 1997. What is a processual analysis. *Scandinavian journal of management*, 13(4), pp.337-348.

Pettigrew, A. M., Woodman, R. W., and Cameron, K. S., 2001. Studying organizational change and development: Challenges for future research. *Academy of management journal*, 44(4), pp.697-713.

Reed, M., 2005. Reflections on the 'realist turn' in organization and management studies. *Journal of Management Studies*, *42*(8), pp.1621-1644.

Rice, A. K., 1969. Individual, group and intergroup processes. *Human relations*, 22(6), pp.565-584.
Rubin, H. J. and Rubin, I. S., 2012. *Qualitative interviewing: The art of hearing data*. London: Sage
Saldaña, J., 2013. *The coding manual for qualitative researchers*. London: Sage.
Saxena, D. and McDonagh, J., 2017. Yet Another 'List'of Critical Success 'Factors' for Enterprise Systems: Review of Empirical Evidence and Suggested Research Directions. In: The UK Academy for Information Systems, *UK Academy of Information Systems Conference 2017*. Oxford, UK, April 2017.
Sayer, A., 2000. *Realism and social science*. London: Sage.
Sayer, A., 2010. *Method in Social Science*, 2nd Ed. London: Routledge.
Simon, H.A., 1982. *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). Cambridge: MIT Press, Massachusetts.
Trist, E., 1981. *The evolution of socio-technical systems*. Occasional Paper, 2. Available from: http://www.lmmiller.com/blog/wp-content/uploads/2013/06/The-Evolution-of-Socio-Technical-Systems-Trist.pdf
Tsang, E.W., 2014. Case studies and generalization in information systems research: A critical realist perspective. *Journal of Strategic Information Systems*, *23*(2), pp.174-186.
Tsang, E.W. and Kwan, K.M., 1999. Replication and theory development in organizational science: A critical realist perspective. *Academy of Management review*, *24*(4), pp.759-780.
Van de Ven, A. H. and Poole, M. S., 1995. Explaining development and change in organizations. *Academy of management review*, 20(3), pp.510-540.
Volkoff, O. and Strong, D. M., 2013. Critical Realism and Affordances: Theorizing IT-Associated Organizational Change Processes. *MIS Quarterly*, *37*(3), pp.819-834.
Walsh, I., Holton, J.A., Bailyn, L., Fernandez, W., Levina, N. and Glaser, B., 2015a. Rejoinder: Moving the management field forward. *Organizational Research Methods*, *18*(4), pp.620-628.
Walsh, I., Holton, J.A., Bailyn, L., Fernandez, W., Levina, N. and Glaser, B., 2015b. What grounded theory is… a critically reflective conversation among scholars. *Organizational Research Methods*, *18*(4), pp.581-599.
Winter, S., Berente, N., Howison, J. and Butler, B., 2014. Beyond the organizational 'container': Conceptualizing 21st century sociotechnical work. *Information and Organization*, 24(4), pp.250-269.
Wolcott, H., 2009. *Writing Up Qualitative Research*. 3rd Ed. London: Sage.
Wynn Jr, D. and Williams, C. K., 2012. Principles for Conducting Critical Realist Case Study Research in Information Systems. *MIS Quarterly*, 36(3), pp.787-810.
Yin, R. K., 2013. *Case study research: Design and methods.* 5th Ed. London: Sage.
Zamutto, R. G., Griffith, T. L., Majchrzak, A., Dougherty, D. J. and Faraj, S., 2007. Information Technology and the Changing Fabric of Organization, *Organization Science.* 18(5), pp.749-762.