

REMINDERS MAY INCREASE RESPONSE RATES, BUT IS THERE A COST?
THE EFFECTS OF SURVEY REMINDERS ON SUBOPTIMAL RESPONSE BEHAVIOR

A Dissertation

Presented to

The Faculty of the Department

of Psychology

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

By

Daniel J. Ingels

May, 2020

REMINDERS MAY INCREASE RESPONSE RATES, BUT IS THERE A COST?
THE EFFECTS OF SURVEY REMINDERS ON SUBOPTIMAL RESPONSE BEHAVIOR

An Abstract of a Dissertation

Presented to

The Faculty of the Department

of Psychology

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

By

Daniel J. Ingels

May, 2020

ABSTRACT

Organizational survey researchers and practitioners must be thoughtful about the methods used to encourage potential survey participants to complete organizational surveys. One of the most common tools used is the survey reminder, which is effective in improving response rates. However, little research has considered whether the responses gained after reminders have been distributed are of comparable quality to those obtained after initial survey invites were sent. Drawing on suboptimal response and nonresponse theories, I examine whether reminders lead survey participants to respond suboptimally, including through insufficient effort response and socially desirable response, as well as if reminders lead to lower survey data quality. Using survey responses from 5,900 respondents to an organizational safety survey, results from measurement and structural invariance tests and concurrent *t*-test analyses show that reminders are not significantly associated with response distortion or measurement variance, indicating that survey reminders do not contribute to lower survey data quality.

TABLE OF CONTENTS

Introduction	1
Increasing Employees' Motivation to Complete a Survey	6
Respondent Behavior after Receiving Survey Reminders	10
Effects of Reminders on Observed Relationships between Measured Constructs	12
Study 1	13
Method	13
Procedures and Participants	13
Measures	14
Analyses	16
Replication of Relevant Meta-Analysis	18
Results	19
Measurement Invariance Test Results	20
Independent Samples t-Test Results for Concurrent Reminders	22
Structural Invariance Test Results	24
Discussion	26
Study 2	26
Method	30
Participants and Procedures	30
Measures	30
Analyses	30
Results	31
Discussion	31

General Discussion	32
Theoretical Implications	32
Practical Implications.....	36
Limitations and Directions for Future Research	38
Conclusion	43
References	44

LIST OF TABLES AND FIGURES

Table 1	53
Table 2	56
Table 3	57
Table 4	58
Table 5	58
Table 6	59
Table 7	60
Table 8	61
Table 9	62
Table 10	62
Figure 1	63
Figure 2	64

Reminders May Increase Response Rates, but Is There a Cost?

The Effects of Survey Reminders on Suboptimal Response Behavior

In organizational science research and practice, survey questionnaires continue to be among the primary methodologies for obtaining data on employees and other research participants. Surveys are uniquely capable of measuring individuals' attitudes, emotions, cognitions, and perceptions at a large scale in organizational contexts in ways that more novel, cutting-edge measurement methods, such as web scraping, gesture analysis, and functional magnetic resonance imaging (fMRI), and other more traditional methods, such as interviews and focus groups, cannot quite as efficiently match. As such, organizational survey consultants, academic researchers, and organizational effectiveness practitioners continue to rely on a variety of different types of surveys to assess attitudes and culture within their organizations or clients. Surveys are relatively affordable, comparatively easy to deploy, able to assess large numbers of individuals, and able to be analyzed through a plethora of quantitative statistical techniques to provide informative and powerful insights into psychosocial phenomena occurring in the work context. Until other measurement methods are more capable of assessing a person's internal characteristics, surveys will continue to have a place in the toolkit for the organizational scientist-practitioner. Since surveys are, and will continue to be, used to measure organizationally-relevant constructs, survey researchers need to use any of a number of strategies to maximize survey response rates. One of the primary strategies is to provide reminder messages to potential participants. A number of studies by prominent survey researchers have shown that sending multiple survey reminders to potential participants can be among the most fruitful methods of increasing response rates, potentially by as much as 20-40 percentage points (Asch,

Jedrziewski, & Christakis, 1997; Dillman, 2007; Kanuk & Berenson, 1975). Other methods for inducing survey response, such as reducing survey length or providing incentives for completion, have been shown to improve response rates, but they are not as effective as survey reminders in improving response rates (Dillman, 2007; Porter, 2004). Thus, although survey researchers have many tools that they can utilize to increase survey response rates, the effectiveness of survey reminders means that survey researchers will continue to use reminders as part of the survey administration process.

This study will consider the effects of survey reminders on response behavior in order to determine if reminders can contribute to measurement problems, testing if response rate-boosting techniques can have potential negative consequences. As such, the empirical goal of this study is to determine if receiving one or more reminders contributes to predictable measurement bias through suboptimal responding, including insufficient effort responding and socially desirable responding.

Surveys are not without meaningful faults in measuring organizational phenomena adequately and accurately. In particular, survey results can be biased through two categories of survey-related behavior: nonresponse and suboptimal response. *Nonresponse* consists of failure to respond to a survey either completely (i.e. unit nonresponse) or to one or more questions (i.e. item nonresponse). Nonresponse can typically be categorized into two domains: *active* and *passive* nonresponse. *Active* nonresponse consists of a potential survey participant knowing and recognizing that they can complete the survey and making an active decision to not complete the survey, while *passive* nonresponse consists of a potential survey participant not actually receiving the survey, forgetting about the survey, missing the survey (e.g. due to missing work), or being unable to complete the survey due to other commitments

(Rogelberg, Conway, Sederburg, Spitzmueller, Aziz, & Knight, 2003; Rogelberg & Stanton, 2007). Nonresponse is particularly harmful when potential survey takers who share common characteristics respond at lower rates than other potential survey takers with different characteristics, as the results of the survey would be more reflective of the latter segment of individuals than all of the potential survey takers, as a whole (Rogelberg & Luong, 1998; Rogelberg & Stanton, 2007). Nonresponse is also problematic in that it reduces research utility in two additional ways. First, nonresponse provides a lower sample size than initially expected (assuming probabilistic sampling), which potentially restricts the available statistical procedures that can be used to evaluate research hypotheses and decreases statistical power to detect statistical relationships between variables measured in the survey (Rogelberg & Stanton, 2007). Second, key stakeholders are often aware that low sample sizes make survey data less useful, potentially reducing the credibility of the survey data and insights gleaned from it (Rogelberg & Stanton, 2007). This can be particularly problematic when attempting to report on individual units where unit sample sizes are particularly small, as one may not even be able to report on such units or propose interventions based on such limited information. Thus, in knowing of these important measurement issues, organizational survey researchers and survey practitioners ought to feel encouraged to maximize response rates and sample sizes. However, as will be tested in the present study, attempts to maximize response rates may, perhaps, somewhat reflect a zero-sum trade-off by encouraging suboptimal survey-taking behavior that replaces simply not responding to a survey.

Potential survey participants will not just simply either provide an optimal response to a survey or not respond at all; rather, some may provide responses that involve shortcutting the cognitive process necessary to optimally respond to survey questions. This behavior is

termed *satisficing* (Krosnick, 1991; Krosnick, Narayan, & Smith, 1996). Tourangeau (1984) describes four stages of the cognitive process that occurs when individuals optimally respond to a survey item. First, respondents read and comprehend the item's meaning. Second, respondents retrieve all relevant information from their memories. Third, respondents integrate that information in the context of the item in order to make a judgment about the item. Fourth, they report their judgments by deciding which response option best reflects their judgments. Krosnick et al. (1991; 1996) define satisficing as merely satisfactorily, rather than optimally, completing any one of these stages of the response process and categorize two types of satisficing: weak and strong. *Weak satisficing* consists of completing each of the four stages but doing so in a non-thorough manner for at least one stage. Weak satisficing can manifest in, for example, less-thoughtfully reading and comprehending an item, superficially accessing memories, making an overly-quick judgment about the item, or not thoroughly considering which response option to choose. *Strong satisficing* involves not engaging in the retrieval and integration/judgment stages entirely. This essentially consists of either fully or only somewhat comprehending the item and, more importantly, selecting the response that the respondent believes will appear to be an appropriate answer. Strong satisficing can be considered akin to *socially desirable responding* (SDR) (Paulhus, 1984; 2002) in that the responses are not accurate self-reports but rather represent socially-aware judgments of what a reasonable response would be. However, SDR represents an underlying construct representing the desire to present positively while strong satisficing can consist of all response behaviors (i.e. positive, negative, and neutral) that fail to be self-reports in regard to the construct addressed by the item. Essentially, SDR can be considered a specific form of strong satisficing. Lastly, although not detailed in the *satisficing* literature, one additional

form of non-optimal survey completion consists of *insufficient effort*, *careless*, or *inattentive responding* (the term *insufficient effort responding*, or IER, will be used in this manuscript henceforth), which can be defined as response to an item with no regard to item content (careless responding; Meade & Craig, 2012) or response with "low or little motivation to comply with survey instructions, correctly interpret item content, and provide accurate responses" (Huang, Curran, Keeney, Poposki, & DeShon, 2012). IER can consist of a variety of response strategies, including random response and nonrandom response (e.g. straight-line response). Weak satisficing, strong satisficing, and IER are all problematic survey response behavior strategies that can bias survey results and, ultimately, lead to conclusions and insights that do not adequately reflect the real, lived experiences of all the individuals who completed the survey.

The purpose of this study is to assess if a common strategy to improve response rates, the survey reminder, is effective in reducing nonresponse and obtaining a greater amount of quality data or if such a strategy only leads to the additional data being lower quality due to a greater occurrence of satisficing or insufficient effort responding by the participants who were drawn in by the survey reminders. This study will attempt to assess this goal through a three-pronged strategy. First, I will attempt to demonstrate that statistical measures of detecting suboptimal response (e.g. Huang, Bowling, Liu, & Li, 2015) will be dependent upon survey reminders, such that whether one needed a reminder will be related to identification as a suboptimal respondent. Second, I will attempt to demonstrate that reminders lead to poorer quality data from the respondents who needed them by assessing the structural validity of their responses (i.e. invariance of measurement compared to those who did not need reminders). Third, I will test whether reminders can affect the external validity

of statistical relationships between survey constructs (i.e. structural invariance compared to those who did not need reminders). Ultimately, the intention behind this study is to provide evidence to support or discourage the usage of reminders, whether at all or beyond a small number of them.

Increasing Employees' Motivation to Complete a Survey

When organizational survey researchers wish to gather opinion, attitude, or other survey data from their employees, among the most important responsibilities involved in the survey process is convincing those employees to actually complete the survey. Survey researchers and practitioners typically do this by tailoring the survey and the information surrounding the survey (e.g. introductory information, messaging, incentives, etc.) in order to conform to their employees' concerns, interests, and desires (Dillman, 2007; Groves & Couper, 1998).

Two psychological theories have been posited by researchers in the public opinion community to address why tailoring survey-related content to the survey participant base should be effective. The first of these is based in social exchange theory, which involves a series of interactions that lead to two entities sharing obligations toward each other (Cropanzano & Mitchell, 2005; Emerson, 1976). A primary theme in social exchange theory is the concept of *reciprocity*, which suggests that there are normative rules in the bilateral relationship where if one entity provides something to the other entity, the other entity ought to repay with an equivalent action or object (Cropanzano & Mitchell, 2005; Gouldner, 1960). In the context of organizational surveying, the social exchange phenomenon can occur in this ideal format:

1. The survey researcher deploys the survey with information detailing how the survey will be used to improve the organization to employees' benefit, potentially advertising material incentives.
2. The participant completes the survey with the understanding that he or she will be rewarded through improvements in the organization, as well as through the material incentive, if offered.
3. The survey researcher then implements survey-data-driven organizational changes and pays incentives to participants who complete the survey.

The second psychological theory on tailored surveying is leverage-saliency theory, which states that the topics of the survey, along with its messaging, must be both meaningful and apparent to the potential respondent (Groves, Singer, & Corning, 2000). For a survey researcher to have *leverage* over a potential participant, the topics of the survey need to be relevant to the respondent's lived experiences. For example, a survey exclusively focused on early childhood care benefits may not be particularly relevant to an older worker who has no children or whose children are adults. Thus, the survey researcher would have little leverage over this person in order to obtain their response. The topics in the survey must also be *salient*, meaning highly apparent or advertised well, in the survey and messaging materials in order to demonstrate that those topics are highly important to organizational sponsors and will likely be acted-upon. Thus, organizational survey researchers and practitioners are encouraged to design surveys and related materials that are meaningful to the maximal number of individuals within the organization, while making those meaningful topics highly salient to participants and potentially decreasing the salience of harmful or controversial topics. Based on these two theories of tailored surveying, public opinion experts have

generally detailed the following methods for improving survey participation (Dillman 2007; Kanuk and Berenson, 1975; Kypri and Gallagher, 2003; Porter, 2004; Robertson, Walkom, & McGettigan, 2005; Schirmer, 2009; Warriner, Goyder, Gjertsen, Hohner, & McSpurren, 1996):

- Carefully designed surveys (e.g. layout, question content, number of questions, etc.);
- Detailing the relevance of the survey;
- Sending a pre-survey letter to potential participants;
- Advertising and providing material incentives;
- Providing multiple points of contact to participants (i.e. reminders).

Among these methods, the most peculiar is the usage of reminders because, by design, they are not used for all potential survey respondents. Rather, they are only used for potential respondents who do not provide responses immediately after the survey researcher deploys the survey. Potential respondents who have not yet responded can be considered to have not done so either because they lack the motivation to do so and/or are unable to do so at the time (e.g. Maier's Performance Formula; Maier, 1955). Survey reminders can be a potentially powerful tool to increase motivation to respond among those who are unmotivated to do so and can encourage those with little time or opportunity to complete a survey to provide greater priority toward the survey, or to "catch" them at a better time. Previous research on correlates of nonresponse show that nonresponse is highly related to motivational and ability-related factors, including conscientiousness and agreeableness (Rogelberg et al., 2003), perceived organizational support (Spitzmueller, Glenn, Barr, Rogelberg, & Daniel, 2006), organizational citizenship behavior (Spitzmueller, Glenn, Sutton, Barr, & Rogelberg, 2007), and overload-based strain (Barr, Spitzmueller, & Stuebing,

2008). Indeed, reminders may have the potency to push some individuals who have lower levels of conscientiousness, agreeableness, citizenship orientation, and perceived organizational support, as well as those highly overloaded, to provide response to an organizational survey. However, this claim has received no attention, to date, in the bodies of literature in the organizational sciences or public opinion. Although reminders may be useful in converting these individuals from nonrespondents to respondents, one must consider that these individuals still have these characteristics that are conducive to nonresponse. Thus, one must consider if these reminders are leading them to complete a survey optimally or merely provide suboptimal responses in order to meet a perceived requirement to complete the survey.

Although survey reminders are generally used to provide a well-intentioned and mild push to consider completing a survey, reminder targets may perceive reminders as a method to induce pressure to complete the survey. This conceptualization of surveys as a form of organizational pressure is based in the concept of compulsory citizenship behavior (CCB) (Vigoda-Gadot, 2006). CCB consists of behaviors that are generally viewed as organizational citizenship behavior (OCB) (e.g. Organ, 1988; Smith, Organ, & Near, 1983) in that they appear to be outside of typical job requirements that help the organization overall but are implied to be necessary and required by a supervisor or management. Research on CCB is still fairly nascent in the organizational science literature, but research on volition in OCB has shown that a potentially abusive or exploitative manager or supervisor can take advantage of an employee by suggesting that the employee is not in a position to refuse management requests, since refusal may mean that the organization could suffer and the employee may be held accountable (Vigoda-Gadot, 2007; Zellars, Tepper, & Duffy, 2002).

Although survey completion may actually not be a very high-stakes task, survey reminders that are frequent or contain high-pressure language may condition employees to believe that survey completion is of such importance that failing to do so would constitute failing the organization more broadly. In turn, this would make survey completion *appear implicitly mandatory*. Thus, although an organizational survey researcher may very well observe increases in survey completion after reminders, he or she cannot be certain that those additional respondents complete the survey under the same perceived conditions as early respondents who needed no reminders. Rather, respondents who needed reminders in order to complete the survey could be experiencing unintended pressure to complete the survey, potentially in such a way that can affect their responses and affect the validity of the survey's results and conclusions.

Respondent Behavior after Receiving Survey Reminders

Because respondents who need reminders in order to respond to a survey are typically either less-motivated and/or less-able to respond immediately than early respondents, respondents who need reminders may be more likely to take cognitive shortcuts (i.e. engage in suboptimal responding behavior) in order to complete the survey. The concept of survey satisficing comes from the broader study of satisficing in human decision-making, in general. Satisficing theory in decision sciences stems from research in psychology on adaptive behavior and in economics on rational behavior. In both domains of research, an individual typically has to make complex choices in order to meet one's own needs and meet the performance requirements of others, while utilizing resources that are scarce, such as time, money, and cognitive capacity (Simon, 1956; Wierzbicki, 1982). In order to satisfy one's own needs and the needs of others, an individual is generally unable to perform at a maximal

level when achieving each need. Rather, the individual is more likely to determine which level of performance is minimally sufficient in order to satisfactorily meet each need and perform at that level to meet needs perceived to be less important while performing more optimally to meet needs perceived to be of greater importance. If one assumes that early survey respondents are more motivated to complete a survey or find it to be of greater importance when compared to respondents who need reminders, then those who need reminders ought to be more likely to consider survey response to be a low-priority task that can be completed through satisficing behavior.

Hypothesis 1: Respondents who require greater numbers of reminders to complete a survey will be more likely to respond to the survey with insufficient effort.

Beyond simply inducing respondents to respond suboptimally, reminders have the potential to create a high-pressure survey environment that encourages respondents to respond in a socially desirable way. Socially desirable responding can be defined as “the tendency to give overly positive self-descriptions” and includes dimensions of self-deception and impression management toward others (Paulhus, 1984; Paulhus 2002; Zerbe & Paulhus, 1987). The inducement effect of reminders on socially desirable responding is based on signaling theory, which suggests that when two parties have access to differing information, one party must determine how to signal its information, and the other party must decide how to interpret that signal (Connelly, Certo, Ireland, & Reutzel, 2011; Spence, 2002). Signaling theory has typically been used in research to describe how an organization advertises or presents publicly in order to gain public favor or recruit quality applicants (e.g. displaying minority race models on a recruitment website in order to recruit minority job candidates). Regarding survey reminders, although a survey researcher may send reminders in order to

simply attain a high volume of survey responses, potential respondents may interpret reminders as an indication that not only is completion of the survey desired but doing so perhaps appears to be a *requirement* for which they are being monitored by the organization. Therefore, potential participants may interpret that monitoring for survey completion may be conflated with monitoring of specific responses, giving those who receive reminders a greater perceived need to manage impressions through their responses to the survey in order to avoid appearing highly critical of the organization or diminishing their own reputations. Thus, for those who receive reminders to complete a survey, there ought to be greater evidence of socially desirable responding.

Hypothesis 2: Respondents who require greater numbers of reminders to complete a survey will be more likely to respond to the survey in a socially desirable manner.

Effects of Reminders on Observed Relationships between Measured Constructs

If the assumption that reminders negatively affect measurement quality holds true, then the relationships between surveyed variables should also be distorted, since those variables would not truly represent the innate levels of the construct for those individuals who receive reminders. In regard to measurement properties of survey scales, suboptimal response can be considered a major source of random error, which can harm a measure's factor structure (Huang et al., 2012) and introduce noise into each measure (Schmitt & Stults, 1985; Woods, 2006). However, some disagreement exists in the literature on suboptimal response behavior on how such response affects the observed structural relationships between variables, with some suggesting that suboptimal responses can attenuate observed relationships between variables (i.e. Type II errors) (McGrath, Mitchell, Kim, & Hough, 2010) and others suggesting that suboptimal responses can inflate observed relationships (i.e.

Type I errors) (Huang, Liu, & Bowling, 2015). Regardless of how suboptimal responses affect the relationships between observed variables, the scientific consensus is that suboptimal responses do distort the observed structural relationships between variables. Thus, one could expect that the relationships between observed variables will be distorted for those who need to be reminded to respond, in comparison to the early responders who do not need reminders.

Hypothesis 3: Assessments of relationships between theoretically related constructs will be distorted for those who require greater numbers of reminders to complete a survey.

STUDY 1

Method

Procedure and Participants

This study used survey data in which participants were recruited through a large organization in the energy industry based in the southwestern United States, with operations throughout North America. The sample collection targeted over 7,000 full-time employees of this organization as part of an assessment of workplace safety for both office workers and field workers, and 5,900 employees responded to the survey, providing a final response rate of 79.8%.

The procedure for collecting survey data occurred via web-based distribution through email, as all employees had regular access to their work email at least each work day. Links were embedded in the emails and were deployed from the client organization's SurveyMonkey account. Responses were tracked through SurveyMonkey's interface and were open for collection over a period of two weeks, with reminders being sent at various

points over those two weeks. Reminders were sent 0-1 times each day in the first week and 1-2 times each day in the second week by a member of the Human Resources department in the client organization, for a total of 14 reminders. Reminders were not sent to participants after they had completed the survey. Given the total number of reminders sent to participants, it is important to recognize potential ethical issues in providing multiple reminders to targeted participants. Schirmer (2009) details several principles for ethical use of survey reminders. We ensured that we adhered to these principles to the highest extent possible, including notifying targeted participants that they would receive reminders, ensuring participation was voluntary, reminding only nonrespondents, providing context and relevance in the reminder messages, and ending reminders after an acceptable response rate was achieved.

Responses were linked to the total number of reminders required before each participant's timestamped completion, and the number of reminders required to complete the survey are used as quasi-experimental conditions for analyses. Participants received no monetary incentive or compensation for completing the survey, and participants were notified through internal communicate that participation was optional.

Measures

All items were assessed using a 5-point Likert-style scale where respondents were asked to indicate their agreement with statements ranging from 1 (strongly disagree) to 5 (strongly agree), with 3 indicating a neutral response. Given that the primary purpose of this manuscript is to evaluate survey-taking conditions, and not further the science of occupational safety, the scales used are summarized in Table 1.

Insert Table 1 about here

In addition to the survey measures, indices of insufficient effort were calculated through methods detailed by Huang et al. (2012). These researchers indicate four methods of measuring insufficient effort: an inconsistency approach, a response pattern approach, an infrequency approach, and a response time approach. The inconsistency and response pattern approaches can be measured by calculating indices based on actual responses to the survey and will be utilized in the analyses for this study. The inconsistency approach assumes that some poorly motivated respondents will be more likely to provide inconsistent responses than highly motivated respondents (Huang et al., 2012). The response pattern (or over-consistency) approach assumes, on the other hand, that some poorly motivated respondents will be more likely to provide overly consistent responses than highly motivated respondents who may indicate more nuanced responses (Huang et al., 2012). The measures to be used for inconsistent responses include Mahalanobis distance, split-half survey inconsistency (e.g. overly low within-person correlations between items in each half of the survey), psychometric synonyms inconsistency (e.g. overly low within-person correlations between items that are highly positively correlated between respondents), and intra-individual response variability (e.g. overly variable within-person responses). The measures to be used for the over-consistency approach include split-half survey over-consistency (e.g. overly high within-person correlations between items in each half of the survey), psychometric antonyms over-consistency (e.g. overly high within-person correlations between items that are highly negatively correlated between respondents), intra-individual response invariability (e.g. low or non-variable within-person responses), and longstring responses (e.g. several consecutive identical responses). All of these IER indices are defined by Yentes and Wilhelm (2018) and are summarized in Table 2. The response time approach assumes that extremely short times

to complete a survey are indications of insufficient effort responding (Huang et al., 2012). Response time is tracked by calculating the difference between survey start and end timestamps as recorded in the survey platform. Although response “start” and “completion” times were tracked for each respondent in SurveyMonkey, the “start” timestamps tracked the first time that a respondent opened the survey, even though some participants may have closed the survey and completed the survey later. Thus, the response times calculated through SurveyMonkey’s start and completion timestamps were highly contaminated and excluded from the analyses in this study. The infrequency approach uses responses to special items for which respondents should all respond in the same way (e.g. “I was born on February 30th”). We did not include any of these items in the survey, as these items were not related to the jobs of those surveyed nor to the content of the survey. Thus, the infrequency approach was not used in this study.

Insert Table 2 about here

Analyses

To assess Hypotheses 1 and 2, that reminders would lead to greater likelihood of insufficient effort or socially desirable responding, respectively, data was analyzed through two strategies. The first strategy was to conduct measurement invariance tests through confirmatory factor analyses in R, through the ‘lavaan’ package (Rosseel, 2012), using the analytical strategy detailed by Kenny (2011), with reminders vs. no reminders being used as grouping variables. To assess Hypothesis 1, the measurement invariance test would have to lead to poor indices of model fit when testing for equivalence in residual variances. Failing to establish strict invariance would indicate support for Hypothesis 1, as the reminders group

ought to contain residual variance due to other factors not related to their specified constructs. Hypothesis 1 will also be assessed in an additional measurement invariance test by loading all items onto a general person factor (e.g. common method variance factor; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) (in addition to their specified factor constructs) and testing for equivalence in factor loadings on the specified constructs, such that the reminder groups ought to have lower factor loadings (indicating that they contain greater variance attributable to factors other than their specified constructs). Failing to establish metric invariance in this method would indicate support for Hypothesis 1. To assess Hypothesis 2, the measurement invariance tests would have to lead to poor indices of model fit when testing for invariance in item means, as the reminder group ought to have higher item means (notably when all items are re-coded positively). Failure to establish scalar invariance with these tests would indicate support for Hypothesis 2.

The second strategy was to conduct *t*-tests between concurrent reminders (e.g. 3 vs. 4 reminders) comparing mean levels of insufficient effort response indices calculated in R via the ‘careless’ package (Yentes & Wilhelm, 2018) and on an indirect measure of social desirability by comparing survey item means (i.e. with higher scores indicating a greater degree of socially desirable responding). Assessments of Cohen’s *d* were calculated to assess effect size. To decrease risk of conducting Type I error(s), I used Bonferroni-corrected alpha values to assess significance of *t*-values. In addition, to limit the number of *t*-tests conducted on survey items, I assessed modification indices for the scalar invariance model described previously to determine which items would be associated with significant reductions in the χ^2 model fit statistic if the intercepts for those items were freed. These items were tested with

this strategy. All insufficient effort response indices generated through the ‘careless’ packaged were used in this analysis.

Replication of Relevant Meta-Analyses

To assess Hypothesis 3, that reminders, and the resulting suboptimal survey responses that results from them, would distort relationships between variables, I conducted a structural invariance test akin to the process described by Kenny (2011). In this process, I would specify a structural equation model for a particular set of relationships between survey variables. I would run the model allowing the regression path parameters to be freely estimated between the reminder and no reminder groups, and then I would run the model with the regression path parameters constrained to be equal. If this structural invariance test showed significantly poorer fit with the second model, then this would indicate support for Hypothesis 3.

In order to assess Hypothesis 3, I specified two models based on a meta-analysis on workplace safety which considered variables similar to the ones assessed in the present study. Specifically, the workplace safety meta-analysis by Christian, Bradley, Wallace, and Burke (2009) considered person and situation factors that affect safety behaviors and safety outcomes. In their meta-analysis, the researchers concluded that distal person-related factors (e.g. safety attitudes and personality) and situation-related factors (e.g. leadership and safety climate) are precursors to more proximal person-related factors, such as safety motivation and knowledge. These proximal factors were then precursors to safety performance, including compliance and participation. The final outcomes considered were accidents and injuries. Many of the variables in the present survey conform to the theoretical model tested by Christian and his colleagues (with the notable exception of personality characteristics), so

this theoretical model will be assessed for structural invariance between those who needed reminders and those who did not need reminders. Two structural models were assessed based on this strategy to ensure whether a less complex model and a more complex model demonstrated invariance. The first model (see Figure 1) involved assessing direct relationships between survey constructs. The second model (see Figure 2) involved loading survey constructs onto higher-order constructs according to the Christian et al. model (i.e. distal person factors, distal situation factors, proximal person factors, safety behavior, and safety outcomes).

Insert Figures 1 and 2 about here

Results

Because of the large size of the tables required, means, standard deviations, and inter-variable correlations for number of reminders, survey construct variables, and insufficient effort responding indices are presented digitally in Supplemental Tables or can be provided upon request by contacting the author of this manuscript: one that presents statistics by items and one that presents statistics by constructs (i.e. using scale means for participants). The distribution of the total sample size by number of reminders received is presented in Table 3.

Insert Table 3 about here

Measurement Invariance Test Results

Hypotheses 1 and 2, that reminders would lead to greater likelihood of insufficient effort or socially desirable responding, respectively, were evaluated through measurement invariance tests that adhere to the process detailed by Kenny (2011) by comparing the model fit differences between nested models. The first model, testing configural invariance, would ensure that the factor structure is invariant for both the reminder group and no-reminder group. The second model, testing metric invariance, constrains factor loadings to be equal between groups and is then compared to the configural invariance model to determine if there is deterioration in model fit. Subsequently, a third model, testing scalar invariance, constrains item intercepts/means to be equal between groups. Lastly, the fourth model, testing strict invariance, constrains item residual variances to be equal between groups. It should be noted that χ^2 tests of model fit and differences in model fit are highly sensitive to large sample sizes and should not necessarily be used as evidence to support or reject study hypotheses (Schlermelleh-Engel et al. 2003; Vandenberg 2006). Indeed, this was the case in the present study. Thus, alternative model fit indices (e.g. CFI, TLI, RMSEA, and SRMR) will be primarily used to assess model fit and differences in model fit for invariance testing. Kline (2005) and Hooper, Coughlan, and Mullen (2008) recommend using the following fit indices to assess model fit for a single model: greater than or equal to 0.90 for CFI, greater than or equal to 0.95 for TLI, less than or equal to 0.08 for RMSEA, and less than 0.08 for SRMR. Chen (2007) recommends using the following differences in alternative fit indices to assess whether there are meaningful differences between nested models: -0.01 for Δ CFI, 0.015 for Δ RMSEA, and 0.030 for Δ SRMR.

The results of the measurement invariance test of the measurement model *not including a general person factor* are presented in Table 4. Likely due to the large sample size in this study, each of the tested nested models demonstrated significant χ^2 model fit values, but the alternative fit indices each indicated good model fit for each model according to Kline and Hooper et al.'s recommendations (TLI was very close to indicating good fit at the strict invariance level). In regard to the differences between nested models, although χ^2 differences between nested models were significant, the differences between alternative indices for comparing nested models were all within Chen's guidelines. Thus, measurement invariance between the reminders group and no reminders group was established. Because model fit did not substantially deteriorate between the models that had freed and constrained item residual variances (Model 3 vs. Model 4), there could not be a claim that other factors (including insufficient effort responding due to reminders) besides the evaluated constructs were responsible for variance in item responses. Thus, Hypothesis 1 was not supported by this analysis. As well, because model fit did not substantially deteriorate between the models that had freed and constrained item intercepts/means (Model 2 vs. Model 3), there could not be a claim that outside factors (including socially desirable responding due to reminders) were responsible for group differences in average response to survey items. Thus, Hypothesis 2 was not supported by this analysis.

Insert Table 4 about here

The results of the measurement invariance test of the measurement model *including a general person factor* are presented in Table 5. Alternative fit indices indicated good model

fit for each model (again, TLI was very close to indicating good fit at the strict invariance level). As well, the differences between alternative indices for comparing nested models were all within Chen's guidelines, and measurement invariance between the reminders group and no reminders group was established. In this case, model fit did not substantially deteriorate between the models that had freed and constrained factor loadings (Model 1 vs. Model 2). Thus, no claim could be made that factor loadings are shifted more in favor of the general person factor among those who received one or more reminders, as constraining the factor loadings to be equal led to a similarly well-fitting model. Thus, Hypothesis 1 was not supported by this analysis. Again, as with the first measurement invariance test, there was no substantial deterioration in model fit between the models that had freed and constrained item intercepts/means (Model 2 vs. Model 3). Thus, Hypothesis 2 was not supported by this analysis.

Insert Table 5 about here

Independent Samples t-Test Results for Concurrent Reminders

Hypotheses 1 and 2 were also tested by conducting independent samples *t*-tests, on means of indices for insufficient effort responding and item means (to indirectly assess socially desirable responding), with concurrent reminders as grouping factors (e.g. comparing mean differences between 3 vs. 4 reminders) by comparing 0 vs. 1 reminder through 13 vs. 14 reminders. Test-wise alpha levels were adjusted using Bonferroni corrections (α / c). For tests on insufficient effort responding indices, 84 tests were conducted (6 IER indices times 14 comparisons), leading to a test-wise alpha level of 0.000595 for an

experiment-wise alpha of 0.05. Results for the concurrent reminder comparison *t*-tests on insufficient effort responding indices are presented in Table 6, where the reported numeric values represent Cohen's *d* effect sizes, and stars represent *p*-values for the associated *t*-tests. All of the comparisons in this test were *nonsignificant*, indicating that there are no differences in indices of insufficient effort responding between concurrent reminders. Thus, Hypothesis 1 was not supported by this analysis, and there is no significant difference in insufficient effort between individuals who received a certain number of reminders and individuals who received one additional reminder.

Insert Table 6 about here

To assess Hypothesis 2, that there are differences in items means (to assess differences in socially desirable responding), the same process as the previous paragraph was used. In order to restrict the number of tests conducted (minimizing the extent to which test-wise alphas are corrected), I only assessed a subset of the items that were surveyed. To choose these items, I evaluated modification indices of the measurement invariance tests conducted in the first part of the analyses. Using the model without a general person factor, I used the 'lavTestScore' function in R's 'lavaan' package on the scalar invariance model to determine which items would significantly improve model fit if their intercepts/means were free to vary between the reminder and no-reminder groups. Although χ^2 tests were not used to assess model fit, I used reduction in χ^2 as the decision tool as it was an objective measure of which items would potentially be most susceptible to item intercept/mean differences due to membership in the reminder vs. no-reminder groups. The cutoff point for items for

inclusion in the t -test analysis was a χ^2 reduction of 3.84 ($df = 1$), which corresponds to a p -value of 0.05 on a chi-square distribution. This led to 26 items being included in the analysis, and these items are presented in Table 7. Thus, for the 364 tests (26 items times 14 comparisons), the test-wise alpha level (for an experiment-wise alpha of 0.05) was 0.000137. Results for concurrent reminder comparison t -tests are presented in Table 8, where the reported numeric values represent Cohen's d effect sizes, and stars represent p -values for the associated t -tests. All of the comparisons in this test were, again, *nonsignificant*, indicating that there are no differences in item means between concurrent reminders. Thus, Hypothesis 2 was not supported by the t -test analysis, suggesting no significant difference in these indices of socially desirable responding between individuals who received a certain number of reminders and those who received one additional reminder.

 Insert Tables 7 and 8 about here

Structural Invariance Test Results

Hypothesis 3, that reminders may distort the relationships between constructs measured through a survey, was evaluated through structural invariance tests that adhere to the process detailed by Kenny (2011) by comparing model fit differences between nested structural equation models in order to assess if needing one or more reminders vs. needing no reminders results in substantial differences in relationships between variables. The first model allows latent construct variances, latent construct covariances, and regression paths to freely vary between groups, while the measurement parameters (factor loadings, item intercepts/means, and residual variances) are constrained to be equal between groups. The

second, third, and fourth nested models constrain latent variances, latent covariances, and regression paths, respectively, to be equal between groups. As with the measurement invariance tests conducted, structural equation model χ^2 tests of model fit and differences in model fit are highly sensitive to large sample sizes, so alternative fit indices were used to assess model fit and differences in model fit. The same cutoffs used for assessing nested models for measurement invariance testing were used to assess the extent to which nested models differ in fit.

The results of the first-order structural model (as visually presented in Figure 1) are presented in Table 9. Due to the large sample size, each of the nested models demonstrated significant χ^2 model fit values but otherwise demonstrated reasonable fit according to Kline and Hooper et al.'s recommendations for alternative fit indices (TLI was very close to indicating good fit). When comparing all of the nested models, the differences in alternative fit indices were all within the guidelines prescribed by Chen. Thus, structural invariance between the reminders group and no reminders group was established, particularly between the models where regression paths were constrained and freed between the reminders and no reminders groups. Therefore, Hypothesis 3 was not supported by this test.

Insert Table 9 about here

The results of the second-order structural model (as visually presented in Figure 2) are presented in Table 11. Each model's fit was significant according to χ^2 model fit values, and alternative fit indices showed that the models had mediocre fit. However, CFI, RMSEA and SRMR were all close to being within the recommended guidelines for each of the

models. When comparing nested models, the differences between nested models in regard to alternative fit indices were again minimal and within the guidelines recommended by Chen. Thus, structural invariance between the reminders and no reminders group was established, particularly between the models that had constrained and freed regression paths between the reminders and no reminders groups. Therefore, again, Hypothesis 3 was not supported by this test.

Insert Table 10 about here

Discussion

Contrary to each of the hypotheses proposed, reminders had no sizable effect on response behavior, either through insufficient effort responding (*H1*) or socially desirable responding (*H2*), measurement quality, or latent construct relationships (*H3*). Although the null findings of this study cannot be used to conclude that there is absolutely no effect of reminders on response behavior and data quality (e.g. null findings interpretation; Kluger & Tikochinsky, 2001), the results do suggest that any effect is minimal. Thus, based on these findings, the positive effects of using survey reminders (e.g. maximizing response rates) are not attenuated by the potential negative response and measurement effects. Further discussion of these findings are presented in the General Discussion in this manuscript.

STUDY 2

Although the goal of this study is to assess the effects of reminders on survey measurement quality, it is necessary to consider how potentially-confounding third variables may lead survey respondents to both need reminders to complete a survey and respond

suboptimally. Specifically, respondents' levels of the personality trait conscientiousness could potentially play a role in both their need for reminders and propensity to respond suboptimally. Conscientiousness, one of the Big Five personality traits (Barrick & Mount, 2009), can generally be described as one's consistency in behavior that is, for example, dependable, industrious, efficient, and achievement-oriented. Individuals who are highly conscientious will tend to act dutifully and reliably in order to meet job performance goals (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001) and are more likely to engage in organizational citizenship behaviors (Borman, Penner, Allen, & Motowidlo, 2001). Indeed, Hertz and Donovan (2000) found that conscientiousness relates to citizenship behavior through dedication to one's job and interpersonal facilitation, such that highly conscientious individuals engage in citizenship behaviors because they are persistent and committed to the goals of their work role and in order to be a team player and cooperate with others in the organization. Thus, given that organizational survey response can be considered a particular form of citizenship behavior, conscientiousness plays a particular role in motivating survey response such that highly conscientious individuals may be more likely to respond because they are dedicated to the organization's success and want to support the people with whom, or for whom, they work. As will be discussed, conscientiousness is associated with survey nonresponse and suboptimal response behavior, but research has yet to explore if respondents systematically differ in regard to conscientiousness based upon whether they need reminders, or do not need reminders, prior to response. This study will evaluate this research gap in order to establish whether the effects of reminders on survey response quality, as assessed in Study 1, are associated with the effects of reminders or are potentially present as a result of

differences in conscientiousness between respondents who did and did not need reminders to respond.

Previous research findings have supported the hypothesis that conscientiousness is associated with various response behaviors, including nonresponse, insufficient effort response, and socially desirable response. First, Rogelberg et al. (2003) did find, through a population profiling study, that conscientiousness, as measured in an initial survey, is positively associated with survey nonresponse in a subsequent survey. Given that researchers and practitioners essentially use reminders to convert both active and passive nonrespondents into respondents, it can be concluded that the impact of reminders on respondents may, to some extent, be related to conscientiousness as those reminders are being sent to potential respondents that are systematically lower in conscientiousness, on average, than those who responded without reminders. Second, Bowling et al. (2016) found, by assessing acquaintance-reported personality traits, that conscientiousness is associated with insufficient effort responding, such that those lower in conscientiousness were more likely to respond with insufficient effort. This relationship exists due to the nature of less-conscientious individuals tending to be less careful, thorough, and deliberate in their survey responses. Third, on the other hand, Ones, Viswesvaran, and Reiss (1996) found in a meta-analysis of the relationships between Big Five traits and socially desirable responding, using both self- and other-rated measures of personality, that highly conscientious individuals were more likely to engage in socially desirable responding than less-conscientious individuals. As the existent research has demonstrated, conscientiousness is highly intertwined with the survey response behaviors examined in Study 1, and the goal of Study 2 is to assess whether individual differences in conscientiousness are related to the need for reminders in order to

respond. Thus, if such a relationship is supported, then the implication would be that the measured effects of reminders on survey response behavior is essentially an indirect indication of the effects of conscientiousness on survey response behavior.

Although conscientiousness can be linked to survey nonresponse and suboptimal response, the link between conscientiousness and responding after one or more reminders may not be as certain. Indeed, individuals who hold lower levels of conscientiousness may receive more reminders as they are less likely to respond to a survey immediately, but this phenomenon does not necessarily equate to the process that occurs after the reminder has occurred. In concept, reminders serve as a renewed call to complete a survey, which suggests that nonrespondents who are higher in conscientiousness (although not as highly conscientious, on average, as the initial respondents who did not need a reminder) will be more likely to respond to the survey after the reminder than those initial nonrespondents who are even lower in conscientiousness. Thus, in theory, survey reminders may not necessarily successfully convert the nonrespondents who are lowest in conscientiousness. In other words, respondents who do not need reminders in order to respond may be higher in conscientiousness than respondents who do need reminders, but the differences in conscientiousness, on average, may be marginal or insubstantial, leading to the assumption that conscientiousness would not have a meaningful relationship with need for reminders and, ultimately, would not be responsible for any differences in response behavior based on one's need for reminders.

Proposition: Are potential respondents who are lower in conscientiousness more likely to need reminders in order to respond to a survey than potential respondents who are higher in conscientiousness?

Method

Participants and Procedures

Survey data from a sample of students at a large state university in the southwestern United States were collected using the population profiling technique also used by Rogelberg et al. (2003), to assess whether conscientiousness assessed in a captive audience was significantly related to need for a survey reminder for a follow-up survey that occurred about 5-7 months after the captive survey. The initial captive audience surveys occurred in different university classrooms over a two-month period, and the follow-up survey was sent to all respondents online at the same time. The follow-up survey resulted in 35 total complete cases (after matching cases from the captive survey and the follow-up survey). Eight participants completed the follow-up survey without a reminder, and 27 participants completed the follow-up survey after receiving a reminder.

Measures

Conscientiousness was assessed in the captive audience survey using Thompson's (2008) International English Big-Five Mini-Markers conscientiousness 8-item subscale. Respondents rated their level of agreement or disagreement with single-word items that may or may not describe themselves, such as "Organized" and "Systematic."

Need for a survey reminder was operationalized through a binary variable that indicated whether a respondent's survey was completed after the reminder was sent to nonrespondents to complete the online follow-up survey.

Analyses

To assess differences in levels of conscientiousness between respondents who needed reminders and those who did not need reminders, an independent samples *t*-test was

conducted. Because there is no theoretical reason to assume that the variances in conscientiousness were equal between the two groups, the Welch-Satterthwaite procedure for estimating standard error for the t -test was used.

Results

The mean level of reported conscientiousness for those who did not need reminders was 4.016, and the mean level of conscientiousness for those who did need reminders was 4.037. Using a pooled standard deviation of 0.682, the resulting Cohen's d effect size was equal to 0.031, indicating a very small effect. There was no significant difference between the no-reminder group and reminder group in conscientiousness ($t = 0.069$, $df = 11$, $p = 0.947$). Thus, since the effect is small and nonsignificant, there is initial evidence that there are minimal differences in levels of conscientiousness between individuals who respond to a survey without needing reminders and those who respond after receiving reminders.

Discussion

The concern about the role of conscientiousness when assessing the effects of reminders on survey data quality is that one's need for reminders and suboptimal response are both possible outcomes of lower levels of conscientiousness. If this proposition was correct, then reminders should not actually be a strong and direct potential contributor to suboptimal response behavior. However, although the null findings from Study 2 are not an indication that conscientiousness plays *no* confounding role, the results do show that conscientiousness does not have a strong, or even moderate, effect on one's need for a reminder to complete a survey. Thus, although Study 1 found no sizable effect of reminders on survey data quality and suboptimal response, if a sizable effect did exist, conscientiousness would not likely play the confounding role proposed.

GENERAL DISCUSSION

Survey research and practice in organizational contexts rely upon sufficiently high response rates in order to conduct complex analyses and provide generalizable, meaningful results, and reminders are among the most highly-utilized tools to maximize response rates. However, only limited prior research in the organizational sciences has examined the potential downsides of utilizing reminders, particularly in regard to measurement quality. This study addressed the potential for reminders to contribute to four issues in survey measurement: (a) increased propensity to respond with insufficient effort, (b) increased propensity to respond in a socially desirable manner, (c) measurement variance between those who need reminders and those who do not need reminders in order to complete a survey, and (d) differences in the relationships between constructs (i.e. structural invariance) between those who need reminders and those who do not need reminders. The results from this study demonstrated, in the context of an organizational safety climate survey in an energy firm, that each of those potential issues in survey response and measurement are fairly resistant to the effects of reminders, such that reminders do not appear to result in suboptimal response or issues in measurement quality. Again, it is vital to clarify that even though the results of this study show that reminders do not have a significant effect on measurement quality or response behavior, the results should not be interpreted as implying that reminders have no effect at all (Kluger & Tikochinsky, 2001).

Theoretical Implications

One purpose of this study was to address whether reminders induce respondents to complete a survey with distorted responses, which, although not directly assessed, would be attributable to intentional or unintentional pressure from the organization. One pattern of

distorted response tested in this study was insufficient effort responding, or response patterns indicating lack of motivation to comply with survey instructions or provide accurate responses. The assumption in this case was that individuals who need reminders are, in general, either not motivated to respond or lack the resources needed to respond (e.g. time) and, therefore, do not respond upon the initial release and notification of the survey. Survey reminders may encourage most potential respondents to respond in a full and faithful manner, but these reminders do encourage potential respondents to respond when they are not necessarily in the best condition to properly respond. However, results for both the measurement invariance tests (i.e. invariant residual variances in the no-general factor measurement model and factor loadings on the person variable in the general factor measurement model) and IER *t*-tests demonstrated that the potential low motivation and/or lack of resources that resulted in initial nonresponse and *requirement of one or more reminders* did not ultimately have a meaningful effect on responses, on average. Although I caution that these results may be specific to the study's context, this study's results do show that reminders do not have measurable effects on respondent's suboptimal response behavior.

The other pattern of distorted response tested in this study was socially desirable response, or response patterns indicating motivation to respond in a manner that reflects adherence to perceived socially-acceptable norms for behavior or attitudes. The proposition in this study was that reminders serve as signaling devices from the organization that signal, from the perspective of the potential respondent, that the organization *expects* response to the survey and *values* the survey topic to such an extent that there are organization-wide expectations regarding what levels of behavior and attitudes, in the context of the survey topic, are required of respondents. As with the results for IER, the tests of invariance of item

intercepts/means in the measurement invariance tests and the *t*-tests on item mean responses demonstrate that the additional pressure potentially produced by survey reminders did not have more than a minimal effect on the responses to survey items, on average. Thus, the study's results show that, at least in the context of the study, survey reminders do not meaningfully induce respondents to respond in a socially desirable manner.

Although the analyses specific to IER and socially desirable responding demonstrated that reminders did not reflect mean differences in indicators of those constructs, the primary goal of this study was to assess whether a survey systematically measures organizationally-relevant psychological constructs *differently* for those who do not need reminders to complete a survey and those who do need reminders. The results of two measurement invariance tests, one with a general person factor and one without a general person factor, both demonstrated strict invariance. Thus, given that factor loadings, item means, and residual variances are invariant between those who did and did not need reminders, the scales used in the organizational survey deployed in this survey can be claimed to measure the same constructs for both groups. Given that the sample size for this study was quite large (nearly 6,000 participants), if the survey did have measurement variance between groups, there ought to have been enough statistical power to detect the measurement variance between the two groups. Therefore, not only do reminders not simply reflect indications of suboptimal response, but the survey used in this context appears to be highly resistant to measurement differences between those who do and do not receive reminders.

As well, the findings are comparable when testing first-order and second-order structural models, based on Christian et al.'s (2009) meta-analysis on the effects of person and situation factors on safety behavior and outcomes, which showed that reminders also do

not distort the relationships between variables that are theoretically linked, as tests of invariance of latent construct variances, latent construct covariances, and latent construct regression paths did not provide support to the hypothesis that the measurement distortion from reminders would result in distortions in the structural relationships between measured constructs. Thus, the results of this study suggest that survey reminders may not cause respondents to respond in suboptimal ways and ultimately may not lead to distorted measurement or analytical results.

In regard to existing research on survey response behavior, this study is among the first to consider the connections between different, well-researched domains, particularly nonresponse, insufficient effort response, and socially desirable responding. Researchers in each of these domains have investigated a variety of important concepts, including detection, covariates, data quality, and organizational ramifications. As research in these areas has matured, a reasonable next step in researching these behaviors is to understand what actions or phenomena lead an individual engaging in one type of response behavior to engage in a different type of response behavior. This study assesses one organizational action, the survey reminder, and how it converts nonrespondents into respondents. Passive nonrespondents can generally be considered potential respondents who have not responded for any number of reasons, such as having forgotten about the survey or having not gotten around to completing the survey (Rogelberg & Stanton, 2007). A reminder provides an impetus to complete the survey, but depending upon the reminder message, the frequency of reminders, and when the reminders are sent in the context of the potential respondents' work, reminders could convert those nonrespondents into suboptimal respondents (i.e. IER or SDR respondents). The findings of this study suggest that reminders do not substantially contribute to converting

nonrespondents into suboptimal respondents, but this study does provide a framework for understanding how former nonrespondents may behave once they do become respondents. Using any number of respondent characteristics or organizational interventions, survey response behavior researchers can assess how certain types of reward systems or pressure tactics drive certain types of suboptimal response, as well as what personal characteristics serve as boundary conditions for engaging in certain suboptimal response behaviors. For example, potential respondents with high levels of the “dark” personality traits (i.e. narcissism, Machiavellianism, and subclinical psychopathy; Paulhus & Williams, 2002) may be more likely to engage in suboptimal response behavior after receiving reminders, perhaps in order to overly-manage impressions (SDR) or spite the survey sponsor (IER). Thus, the present study provides a meaningful integration of these various bodies of research that furthers our understanding of survey response behavior.

Practical Implications

Considering that this study demonstrates that survey reminders do not appear to meaningfully contribute to measurement issues for an organizational survey, the primary recommendation for survey practitioners and researchers that arises from this study is that reminders are indeed useful tools to maximize response rates. The concern with providing numerous reminders was that a trade-off may exist between maximizing response rates and maximizing survey response quality. This study demonstrates that this trade-off does not exist. Thus, I would encourage practitioners and researchers to use reminders, in a thoughtful manner, in order to generate large sample sizes, ultimately allowing them to avoid the consequences of extensive nonresponse. First, effective reminders will allow practitioners and researchers to use analytical techniques that require a high degree of statistical power,

such as structural equation models. This may not have been a particularly problematic issue in the sample utilized in this study, as the size of the sample prior to reminders being deployed was very large ($N = 1,581$). However, smaller organizations may need to receive responses from as much of the target sample as possible in order to achieve a sample size that allows similar complex analyses. Second, by allowing survey researchers and practitioners to maximize response rates through reminders allows survey results to sufficiently reflect the targeted population, especially if there are particular jobs that require specific tasks or operate within certain work environments that limit potential respondents' ability to complete a survey upon the initial notification of the survey. Thus, using reminders to maximize response rates will allow survey researchers and practitioners to generate conclusions that are representative of the entire organization, or targeted portion of the organization, and ultimately, make claims that organizational stakeholders will find credible.

From a more technical perspective, the results of the measurement and structural invariance tests conducted allow survey researchers and practitioners to make one-to-one comparisons on survey items between the segments of a sample that do and do not need reminders (Brown, 2015). Specifically, equivalence in factor structure (i.e. configural invariance) means that practitioners and researchers can confirm items are representing the same constructs regardless of whether a respondent needs reminders or not. Equivalence in factor loadings (i.e. metric invariance) means that each item's variance is accounted for by its respective factors in the same way, regardless of one's need for reminders. In other words, constructs assessed in the survey similarly affect how respondents provide responses to the survey items. Equivalence in item intercepts/means (i.e. scalar invariance) means that respondents from both groups effectively have the same mean response to the survey items,

allowing practitioners and researchers to use analytical techniques that consider group mean comparisons. Equivalence in item residual variances (i.e. strict invariance) means that there are no other constructs that are differentially affecting responses to survey items between the reminders and no reminders groups. Thus, by knowing that the survey scales being utilized are *actually measuring the same phenomena*, regardless of group, and are equivalently contaminated by other factors, researchers and practitioners will not have to be concerned that item-specific descriptive statistics are affected by the effects of reminders on respondents' survey responses. As well, the structural invariance tests conducted ensure that the variances in latent constructs and relationships between latent constructs (i.e. latent covariances and regression paths) are equivalent between those who do and do not need reminders, and this allows practitioners and researchers to confirm that results from inferential statistical tests relating two or more constructs are not negatively affected by the effects of reminders. Thus, considering that systematic nonresponse can bias results in favor of groups that are more likely to respond to a survey, utilizing reminders can reduce this bias without introducing any new measurement bias to survey results as a consequence of inducing participants to respond.

Limitations and Directions for Future Research

In this study, we build on prior research on nonresponse, socially desirable responding, and insufficient effort responding. Although we address the relationships between survey response inducement strategies and survey measurement quality, there are gaps in knowledge of survey response behavior that could not be addressed in this study that ought to be assessed in future research. First, the survey reported in this study was conducted with only one organization, and although there is not a great deal of research on surveying

culture within organizations, other organizations' survey practices may lead to context-specific results depending upon the content of the survey administered and the techniques used to facilitate response. The topic of this survey was organizational safety, which may be a topic that is highly important to an organization's employees, especially in an industry with substantial safety risk. Thus, even if response to the survey is not immediate (warranting one or more reminders to respond), respondents may be disproportionately more likely to respond faithfully to such a survey on an important topic than to a survey on a topic that is less interesting to potential respondents. As well, web-based surveys have proliferated over the last few decades, making it incredibly easy for survey practitioners to create and deploy many surveys on many different topics. As organizations begin to feel encouraged to administer more surveys for employee sensing, employees can become overwhelmed by surveys, leading to *survey fatigue* (Weiner & Dalessio, 2006). Extensive surveying efforts can potentially lead to a great deal of overload, resulting in nonresponse or satisficing behaviors on surveys. This can be amplified if potential participants are already experiencing a great deal of cognitive load with work demands that are traditionally job-related and unrelated to the survey administration effort. Thus, future research ought to replicate the results of this study on samples with different survey cultures, in different industries, and on survey content that may be of greater or less interest to target respondents.

Another area for further research would consider how insufficient effort and socially desirable responding are assessed or indicated. This study used a fairly well-accepted set of computational indices (Yentes & Wilhelm, 2018) to provide the likelihood that a respondent engaged in IER that is theoretically supported by IER research (e.g. Huang, Bowling, Liu, & Li 2015; Huang, Curran, Keeney, Poposki, & DeShon, 2012), and socially desirable

responding was assessed by assessing group differences in mean item ratings. These indices are highly indirect in assessing their respective response behavior constructs and may be affected by external contaminants. Thus, future research should consider using check questions as more direct measures of IER and SDR to identify suboptimal respondents. For example, Huang, Bowling, Liu, & Li (2015) developed an infrequency scale of items for which most, if not all, attentive respondents would endorse the “correct” response. An example item from their scale is: “I will be punished for meeting the requirements of my job.” A socially desirable responding scale, such as Paulhus’s (1984) *Balanced Inventory of Desirable Responding*, provides a series of items which individuals who respond in a socially desirable manner would more frequently endorse, such as “I always apologize to others for my mistakes,” and would less frequently endorse, such as “Once in a while I laugh at a dirty joke.” Although using these check items do take up valuable space within a survey, they do provide more concrete assessments of response distortion than the calculated indices and mean differences used in the present study. Thus, future research on survey response behavior and inducements should consider both methods.

As indicated at the end of the introduction to this manuscript, the personality trait of conscientiousness was not assessed as part of the main study but could be a highly important covariate of the need for survey reminders. Given that conscientious is predictive of passive nonresponse, and the purpose of reminders is to mitigate nonresponse, further research must evaluate the multivariate relationships between personality, survey response behavior, and the effects of response inducement. One method to directly assess the relationship between conscientiousness and need for reminders is by using the population profiling method (e.g. Rogelberg et al., 2003) by using conscientiousness scores on pre-hire personality assessments

(assuming the organization assesses personality pre-hire, owns and retains that data, and is able to use it for post-hire, non-selection purposes) as predictive covariates of need for reminders or other inducements. This would avoid the usage of less-generalizable undergraduate samples and provide a more ecologically-valid framework for understanding the nuances of survey response and inducement. However, as pre-hire personality assessments are at risk of faking (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006) and result in range restriction for constructs used for selection purposes (Sackett & Yang, 2000; Schmidt, Oh, & Le, 2006), this method brings its own challenges as well. Thus, further research on personality traits and behavioral tendencies should be pursued to assess the extent to which individuals with certain characteristics are both more likely to need reminders to respond and more likely to respond suboptimally.

The present study used a quasi-experimental method in deploying reminders by allowing a member of the partner organization to send reminders to nonrespondents at his/her own discretion when the rate of completion of the survey was slowing. To further the aims of this study, future research efforts ought to utilize a planned experimental design for providing reminders. For example, one experimental design could be to send reminders at consistent intervals to a randomly-assigned group of nonrespondents and not send reminders to the remaining nonrespondents, or perhaps half of nonrespondents can be provided reminders with one particular message (e.g. indicating the positives associated with response) while the other half are provided reminders with a different message (e.g. indicating the negatives associated with nonresponse) in order to determine which types of reminders or messages are most effective at maximizing response rates while minimizing measurement issues. By systematically testing reminder frequency and messages, researchers can provide much more

scientifically sound advice on the best reminder schemes and strategies in order to effectively encourage nonrespondents to complete an organizational survey in an optimal manner.

Lastly, the present study assessed the effects of one form of survey response inducement tactic, but future research should consider how other types of inducements affect survey response quality. Besides reminders, another common response inducement strategy is the response incentive. Incentives to respond can take a variety of different forms: promised versus prepaid, raffle versus individual incentives, monetary versus non-monetary, etc. Research on the effects of incentives on survey response quality should consider whether an incentive encourages respondents to respond faithfully or whether the value of the incentive results in respondents simply responding in order to complete the survey and earn the incentive. Thus, if the latter phenomenon is occurring, respondents may be more likely to engage in IER in order to complete the survey quickly (when compared to those in a situation without an incentive). However, because an incentive does not necessarily inadvertently portray pressure to respond in a particular manner in the same way that I argue reminders might, incentives ought to be less likely to induce socially desirable response. As well, future research should consider how the variety of different inducement strategies affect various qualities or characteristics of the sample to ensure that these strategies are resulting in a sample that accurately reflects the diversity of the organization's workforce. For example, one meaningful research question could consider whether reminders or other inducements draw in more demographically diverse respondents that would otherwise potentially not exist in the sample if those inducement strategies were not utilized.

CONCLUSION

The goal of this study was to provide evidence to support or caution against the use of reminders to maximize organizational survey response rates. More specifically, I sought to demonstrate whether receiving reminders led to response distortion through insufficient effort responding and/or socially desirable responding, ultimately resulting in measurement variance between those respondents who needed reminders to respond and those who did not need reminders in order to respond. The results of the study demonstrate that reminders do not actually result in the measurement issues hypothesized. Thus, although further research is needed to understand the effects of survey reminders, this study supports the recommendation to survey researchers and practitioners that they can continue to thoughtfully use reminders to maximize response rates and develop stronger conclusions from organizational survey data.

REFERENCES

- Asch, D. A., Jedrzejewski, M. K., & Christakis, N. A. (1997). Response rates to mail surveys published in medical journals. *Journal of Clinical Epidemiology*, 50, 1129–1136.
- Barling, J., Loughlin, C., & Kelloway, E. K. (2002). Development and test of a model linking safety-specific transformational leadership and occupational safety. *Journal of Applied Psychology*, 87, 488-496.
- Barr, C. D., Spitzmueller, C., & Stuebing, K. K. (2008). Too stressed out to participate? Examining the relation between stressors and survey response behavior. *Journal of Occupational Health Psychology*, 13, 232-243.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barrick, M. R., & Mount, M. K. (2009). Select on Conscientiousness and Emotional Stability. In E. A. Locke (Ed.), *Handbook of Principles of Organizational Behavior*, 2nd ed. (pp. 19-39), Chichester, UK: John Wiley & Sons, Ltd.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next?. *International Journal of Selection and Assessment*, 9, 9-30.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317-335.
- Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, 9, 52-69.

- Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, *111*, 218-229.
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, *84*, 822-848.
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*, 2nd ed. New York, NY: The Guilford Press.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464-504.
- Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, *94*, 1103-1127.
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*, 39-67.
- Cropanzano, R., & Mitchell, M. S. (2005). Social exchange theory: An interdisciplinary review. *Journal of Management*, *31*, 874-900.
- Dillman, D. A. (2007). *Mail and Internet Surveys: The Tailored Design Method*, 2nd ed. Hoboken, NJ: John Wiley & Sons Inc.
- Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, *44*, 350-383.
- Emerson, R. M. (1976). Social exchange theory. *Annual Review of Sociology*, *2*, 335-362.

- Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25, 161-178.
- Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology*, 5, 347-358.
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys (1st edition)*. New York: Wiley-Interscience.
- Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-Saliency Theory of Survey Participation: Description and an Illustration. *The Public Opinion Quarterly*, 64, 299-308.
- Hayase, L. K. T. (2009). Internal communication in organizations and employee engagement. UNLV Theses/Dissertations/Professional Papers/Capstones. Paper 1176.
- Henning, J. B., Stuft, C. J., Payne, S. C., Bergman, M. E., Mannan, M. S., & Keren, N. (2009). The influence of individual differences on organizational safety attitudes. *Safety Science*, 47, 337-345.
- Hofmann, D. A., & Mark, B. (2006). An investigation of the relationship between safety climate and medication errors as well as other nurse and patient outcomes. *Personnel Psychology*, 59, 847-869.
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6, 53-60.

- Huang, Jason L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology, 30*, 299-311.
- Huang, Jason L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*, 99-114.
- Huang, Jason L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: examining an insidious confound in survey data. *Journal of Applied Psychology, 100*, 828-845.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The “Big Five” revisited. *Journal of Applied Psychology, 85*, 869-879.
- Ivancevich, J. M., & Matteson, M. T. (1980). *Stress and Work: A Managerial Perspective*. Dallas: Scott Foresman.
- Kanuk, L., & Berenson, C. (1975). Mail surveys and response rates: A literature review. *Journal of Marketing Research, 12*, 440-453.
- Kelloway, E. K., Mullen, J., & Francis, L. (2006). Divergent effects of transformational and passive leadership on employee safety. *Journal of Occupational Health Psychology, 11*, 76-86.
- Kenny, D. A. (2011). Multiple groups. Retrieved from <http://davidakenny.net/cm/mgroups.htm>
- Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York, NY: Guilford Press.

- Kluger, A. N., & Tikochinsky, J. (2001). The error of accepting the “theoretical” null hypothesis: The rise, fall, and resurrection of commonsense hypotheses in psychology. *Psychological Bulletin*, 127, 408-423.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.
- Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation*, 1996, 29-44.
- Kuorinka, I., Jonsson, B., Kilbom, A., Vinterberg, H., Biering-Sorensen, F., Andersson, G., & Jorgensen, K. (1987). Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Applied Ergonomics*, 18, 233-237
- Kypri, K., & Gallagher, S. J. (2003). Incentives to increase participation in an internet survey of alcohol use: A controlled experiment. *Alcohol and Alcoholism*, 38, 437-441.
- Maier, N. R. F. (1955). *Psychology in Industry*, 2nd ed. Oxford, England: Houghton Mifflin.
- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, 136, 450-470.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17, 437-755.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660-679.
- Organ, D. W. (1988). *Organizational Citizenship Behavior: The Good Soldier Syndrome*. Lexington, MA: Lexington Books.

- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The Roles of Constructs in Psychological and Educational Measurement* (pp. 49-69). Mahwah, NJ: Lawrence Erlbaum.
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36, 556-563.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879-903.
- Porter, S. R. (2004). Raising response rates: What works? *New Directions for Institutional Research*, 121, 5-21.
- Robertson, J., Walkom, E. J., & McGettigan, P. (2005). Response rates and representativeness: a lottery incentive improves physician survey return rates. *Pharmacoepidemiology and Drug Safety*, 14, 571-577.
- Rogelberg, S., Conway, J. M., Sederburg, M. E., Spitzmueller, C., Aziz, S., & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology*, 88, 1104-1114.
- Rogelberg, S. G., & Luong, A. (1998). Nonresponse to mailed surveys: A review and guide. *Current Directions in Psychological Science*, 7, 60-65.
- Rogelberg, S. G., & Stanton, J. M. (2007). Introduction: Understanding and dealing with organizational survey nonresponse. *Organizational Research Methods*, 10, 195-209.

- Rosseel, Y. (2012). "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software*, 48, 1–36.
- Roth, A. V., Schroeder, R. G., Kristal, M. M., & Huang, X. (2008). *Handbook of Metrics for Research in Operations Management*. Thousand Oaks, CA: Sage Publications.
- Sackett, P. R. & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, 85, 112-118.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23-74.
- Schirmer, J. (2009). Ethical issues in the use of multiple survey reminders. *Journal of Academic Ethics*, 7, 125-139.
- Schmidt, F. L., Oh, I. S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research results. *Personnel Psychology*, 59, 28-305.
- Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, 10, 1-22.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129-138.
- Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, 68, 653-663.
- Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American Economic Review*, 92, 434-459.

- Spitzmueller, C., Glenn, D. M., Barr, C. D., Rogelberg, S. G., & Daniel, P. (2006). "If you treat me right, I reciprocate": Examining the role of exchange in organizational survey response. *Journal of Organizational Behavior*, 27, 19-35.
- Spitzmueller, C., Glenn, D. M., Sutton, M. M., Barr, C. D., & Rogelberg, S. G. (2007). Survey nonrespondents as bad soldiers: Examining the relationship between organizational citizenship and survey response behavior. *International Journal of Selection and Assessment*, 15, 449-459.
- Thompson, E. R. (2008). Development and validation of an International English Big-Five Mini-Markers. *Personality and Individual Differences*, 45, 542-548.
- Tourangeau, R. (1984). Cognitive sciences and survey methods. *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*, 15, 73-100.
- Tucker, S., & Turner, N. (2011). Young worker safety behaviors: Development and validation of measures. *Accident Analysis & Prevention*, 43, 165-175.
- Vandenberg, R. J. (2006). Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? *Organizational Research Methods*, 9, 194-201.
- Vigoda-Gadot, E. (2006). Compulsory citizenship behavior: Theorizing some dark sides of the good soldier syndrome in organizations. *Journal for the Theory of Social Behaviour*, 36, 77-93.
- Vigoda-Gadot, E. (2007). Redrawing the boundaries of OCB? An empirical examination of compulsory extra-role behavior in the workplace. *Journal of Business and Psychology*, 21, 377-405.

- Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & McSpurren, K. (1996). Charities, no; Lotteries, no; Cash, yes: Main effects and interactions in a Canadian incentives experiment. *Public Opinion Quarterly*, 60, 542-562.
- Wierzbicki, A. P. (1982). A mathematical basis for satisficing decision making. *Mathematical Modelling*, 3, 391-405.
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28, 186-191.
- Yentes, R., & Wilhelm, F. (2018). *Procedures for Computing Indices of Careless Responding*. Retrieved from <https://github.com/ryentes/careless/>
- Zellars, K. L., Tepper, B. J., & Duffy, M. K. (2002). Abusive supervision and subordinates' organizational citizenship behavior. *Journal of Applied Psychology*, 87, 1068-1076.
- Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review*, 12, 250-264.
- Zohar, D., & Luria, G. (2005). A multilevel model of safety climate: Cross-level relationships between organization and group-level climates. *Journal of Applied Psychology*, 616-628.

Table 1*List of Scales Used in Survey*

Scale Name	Sample Item	Number of Items	Citation
Organizational Safety Culture	"My organization considers a person's safety behavior when moving/promoting people."	7 items	Zohar and Luria (2005)
Supervisor Safety Culture	"My supervisor makes sure we follow all the safety rules (not just the most important ones)."	9 items	Zohar and Luria (2005)
Individual Safety Culture	"I adhere to safety procedures required by my company."	4 items	Zohar and Luria (2005)
Error Climate	"Employees often think about how they could have prevented mistakes that occur."	4 items	Hofmann and Mark (2006)
Organizational Rewards for Safety	"Workers are rewarded for safety improvement."	3 items	Roth, Schroeder, Kristal, & Huang (2008)
Upward Organizational Communication	"I can expect that safety recommendations I make will be heard and considered."	4 items	Hayase (2009)
Effective Organizational Communication	"I am notified in advance of changes that affect my job safety."	3 items	Hayase (2009)
Open Organizational Communication	"People freely exchange information and opinions."	3 items	Hayase (2009)
Positive Supervisor Communication	"I feel safe telling my supervisor what I am really thinking."	3 items	Hayase (2009)

Table 1 (cont.)*List of Scales Used in Survey*

Scale Name	Sample Item	Number of Items	Citation
Informational Justice	“My supervisor is candid in his/her communications with me.”	4 items	Hayase (2009)
Safety-Specific Transformational Leadership	“My supervisor behaves in a way that displays a commitment to a safe workplace.”	3 items	Kelloway, Mullen, & Francis (2006)
Safety Motivation	“I believe that it is worthwhile to put extra effort into maintaining safety.”	3 items	Griffin & Neal (2000)
Safety Priority	“Sometimes it is necessary to take safety risks to get a job done efficiently.” (reverse-coded)	3 items	Henning, Stuft, Payne, Bergman, Mannan, and Keren (2009)
Safety Voice	“I feel able tell my supervisor about the consequences of dangerous working conditions.”	3 items	Tucker & Turner (2011)
Psychological Safety	“It is safe for me to make suggestions.”	3 items	Edmondson (1999)
Safety Knowledge	“I know how to perform my job in a safe manner.”	4 items	Griffin & Neal (2000)
Safety Attitude	“Safety specific jobs should always get done.”	3 items	Henning, Stuft, Payne, Bergman, Mannan, and Keren (2009)

Table 1 (cont.)*List of Scales Used in Survey*

Scale Name	Sample Item	Number of Items	Citation
Physical Work Environment Safety	“My physical work environment is safe.”	3 items	Griffin & Neal (2000)
Quantitative Overload	“The amount of work I am expected to do is too great.”	3 items	Ivancevich and Matteson (1980)
Qualitative Overload	“My assigned tasks are sometimes too difficult and/or complex.”	3 items	Ivancevich and Matteson (1980)
Mindful Attention and Awareness	“I find myself doing things without paying attention.” (reverse-coded)	3 items	Brown & Ryan (2003)
Safety Behavior - Compliance	“I use the correct safety procedures for carrying out my job at all times.”	3 items	Griffin & Neal (2000)
Safety Behavior - Participation	“I voluntarily carry out tasks or activities that help to improve workplace safety.”	4 items	Griffin & Neal (2000)
Accidents [†]	“Please indicate how frequently you have experienced the following accident(s) in the last six months: caught in, under, or between machines/equipment.”	6 items	Barling, Loughlin, and Kelloway (2002)
Work-Related Musculoskeletal Disorders (WMSDs) [†]	“Please indicate how frequently you have experienced injuries to the following body part(s) in the last six months: wrists/hands.”	7 items	Kuorinka, Jonsson, Kilbom, Vinterberg, Biering-Sorenson, Andersson, and Jorgensen (1987)

[†]Response scales for these items were 7-point frequency scales ranging from 1 (never) to 7 (frequently).

Table 2*Insufficient Effort Responding (IER) Indices*

Index	Description	Type of Indicator
Mahalanobis Distance	The distance of one respondent's data relative to the center of the multivariate distribution of responses.	High values indicate possible Inconsistency IER.
Even-Odd Consistency Index	Divides questionnaire using an even-odd item split and computes the average response across items in each split scale. Then a within-person correlation is computed based on the two split scale scores.	Low values indicate possible Inconsistency IER. High values indicate possible Over-Consistency IER.
Psychometric Synonyms Index	Within-person correlations are calculated for item pairs that are highly positively correlated between respondents.	Highly negative values indicate possible Inconsistency IER.
Psychometric Antonyms Index	Within-person correlations are calculated for item pairs that are highly negatively correlated between respondents.	Highly positive values indicate possible Over-Consistency IER.
Intra-Individual Response Variability	The standard deviation of a respondent's responses to all items within the survey.	High values indicate possible Inconsistency IER. Low values indicate possible Over-Consistency IER.
Longstring Response	The longest length of consecutive identical responses given to items within the survey.	High values indicate possible Over-Consistency IER.

Information in table summarized from Yentes & Wilhelm (2018).

Table 3*Frequencies and Percentages of Survey Reminders*

<i>Reminders Received</i>	<i>N</i>	<i>Percentage of Sample</i>
0	1581	26.80%
1	974	16.51%
2	983	16.66%
3	616	10.44%
4	366	6.20%
5	312	5.29%
6	336	5.69%
7	67	1.14%
8	90	1.53%
9	102	1.73%
10	83	1.41%
11	53	0.90%
12	127	2.15%
13	61	1.03%
14	149	2.53%

N = 5900.

Table 4*Measurement Invariance Test Results for No Reminder vs. One or More Reminders*

	χ^2	<i>df</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>	<i>SRMR</i>	<i>AIC</i>	<i>BIC</i>	$\Delta\chi^2^\dagger$	Δdf^\dagger
Model 1: Configural Invariance	43584.78***	8710	0.926	0.919	0.037	0.062	925743.7	933682.7	-	-
Model 2: Metric Invariance	43794.99***	8783	0.926	0.919	0.037	0.062	925807.9	933259.1	210.21***	73
Model 3: Scalar Invariance	44017.09***	8856	0.925	0.920	0.037	0.062	925884.0	932847.4	222.10***	73
Model 4: Strict Invariance	45072.22***	8954	0.923	0.918	0.037	0.062	926743.1	933051.6	1055.13***	98

N = 5900; *** *p* < 0.001; \dagger Statistics compare model fit to the prior nested model.**Table 5***Measurement Invariance Test Results for No Reminder vs. One or More Reminders, with a General Person Factor*

	χ^2	<i>df</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>	<i>SRMR</i>	<i>AIC</i>	<i>BIC</i>	$\Delta\chi^2^\dagger$	Δdf^\dagger
Model 1: Configural Invariance	36111.90***	8464	0.941	0.934	0.033	0.028	918762.8	928345.8	-	-
Model 2: Metric Invariance	36418.93***	8634	0.941	0.935	0.033	0.029	918729.8	927176.8	307.03***	170
Model 3: Scalar Invariance	36591.66***	8706	0.941	0.935	0.033	0.029	918758.6	926724.4	172.73***	72
Model 4: Strict Invariance	37662.27***	8804	0.939	0.934	0.033	0.030	919633.2	926944.1	1070.61***	98

N = 5900; *** *p* < 0.001; \dagger Statistics compare model fit to the prior nested model.

Table 6*Cohen's d & T-Test Results for Comparisons of Mean Item Responses Between Concurrent Reminders*

<i>Item[†]</i>	<i>Reminder Comparisons</i>													
	<i>0 vs. 1</i>	<i>1 vs. 2</i>	<i>2 vs. 3</i>	<i>3 vs. 4</i>	<i>4 vs. 5</i>	<i>5 vs. 6</i>	<i>6 vs. 7</i>	<i>7 vs. 8</i>	<i>8 vs. 9</i>	<i>9 vs. 10</i>	<i>10 vs. 11</i>	<i>11 vs. 12</i>	<i>12 vs. 13</i>	<i>13 vs. 14</i>
Mahalanobis Distance	0.052	0.144	-0.019	-0.043	0.065	-0.039	0.058	-0.074	-0.035	-0.039	0.264	-0.174	0.087	-0.189
Even-Odd Consistency	-0.029	-0.064	0.000	0.016	-0.104	0.089	-0.044	-0.084	0.053	0.138	-0.103	-0.058	0.072	0.032
Psychometric Synonyms	-0.030	-0.108	0.038	0.022	-0.087	0.010	0.059	0.012	-0.031	-0.004	0.086	-0.116	-0.068	0.151
Psychometric Antonyms	-0.056	0.011	-0.024	-0.002	-0.062	0.024	-0.128	0.062	-0.049	0.177	-0.134	0.021	-0.039	0.087
Intra-Individual Response Variability	0.124	0.013	0.054	-0.061	0.154	-0.074	0.033	0.139	-0.083	-0.109	0.342	-0.265	-0.072	0.019
Longstring Response	-0.084	-0.120	-0.042	0.052	-0.255	0.076	-0.070	0.019	0.039	-0.088	0.025	0.005	0.083	0.075

p-value cut-offs reflect test-wise alpha level adjustments using Bonferroni corrections (α / c), given 84 tests: * $p < 0.000595$ (0.05 / 84), ** $p < 0.000119$ (0.01 / 84), *** $p < 0.0000119$ (0.001 / 84).

Values reported are Cohen's d and stars (*) indicate that the difference in item mean responses is significant according to an independent samples t-test. No tests reached significance.

Table 7*Significant Modification Indices for Item Intercepts in Scalar Invariance Model Tested in No-General Factor Model*

<i>Item</i>	<i>Construct</i>	$\Delta\chi^2^\dagger$
1. "My supervisor is candid in his/her communications with me."	Informational Justice	21.42***
2. "My supervisor frequently talks about safety issues throughout the work week."	Supervisor Safety Culture	18.25***
3. "I never seem to have enough time to get everything done at work."	Quantitative Overload	18.19***
4. "There are significant dangers inherent in the workplace."	Physical Work Env. Safety	14.26***
5. "There are no significant ramifications for skipping safety related meetings."	Safety Behavior - Participation	11.51***
6. "My supervisor makes sure we follow all the safety rules (not just the most important ones)."	Supervisor Safety Culture	11.31***
7. "It is easy to report safety incidents using [reporting system]."	Upward Org. Communication	9.41**
8. "I do not have enough time to get the job done well."	Quantitative Overload	9.16**
9. "How frequently have you experienced the following accidents in the past six months? Open wound, cut, puncture, or infection of the wound."	Accidents	8.53**
10. "I am kept informed about how well safety goals and objectives are being met. "	Effective Org. Communication	8.33**
11. "I am notified in advance of changes that affect my job safety."	Effective Org. Communication	8.17**
12. "It is advantageous to openly discuss one's mistakes."	Error Climate	8.10**
13. "My organization considers a person's safety behavior when moving/promoting people."	Org. Safety Culture	6.90**
14. "Employees are frequently exposed to risky situations."	Physical Work Env. Safety	6.69**
15. "I have insufficient training and/or experience to discharge my duties properly."	Qualitative Overload	6.68**
16. "Sometimes it is necessary to take safety risks to get a job done efficiently." (reverse-coded)	Safety Priority	5.86*
17. "My organization provides detailed safety reports to workers (e.g., injuries, near accidents)."	Org. Safety Culture	5.69*
18. "I feel that it is worthwhile to volunteer for safety related tasks."	Safety Motivation	5.22*
19. "My supervisor offers positive reinforcement to workers who pay special attention to safety. "	Supervisor Safety Culture	5.03*
20. "It is safe to give my opinions."	Psychological Safety	4.89*
21. "The performance standards on my job are too high."	Qualitative Overload	4.57*
22. "Tell my supervisor about hazardous work."	Safety Voice	4.45*
23. "Co-workers do not think it's a big deal to miss safety related meetings."	Safety Behavior - Participation	4.43*
24. "I carry out my work in a safe manner at all times. "	Safety Behavior - Compliance	4.37*
25. "My supervisor makes sure we receive all the equipment needed to do the job safely."	Supervisor Safety Culture	4.00*
26. "Group together with coworkers and take safety concerns to the supervisor."	Safety Voice	3.96*

$N = 5900$; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

† Reduction in scalar invariance model χ^2 if the intercept associated with the respective item constrained in models between groups is allowed to freely vary between group.

Table 8*Cohen's d & T-Test Results for Comparisons of Mean Item Responses Between Concurrent Reminders*

<i>Item[†]</i>	<i>Reminder Comparisons</i>													
	<i>0 vs. 1</i>	<i>1 vs. 2</i>	<i>2 vs. 3</i>	<i>3 vs. 4</i>	<i>4 vs. 5</i>	<i>5 vs. 6</i>	<i>6 vs. 7</i>	<i>7 vs. 8</i>	<i>8 vs. 9</i>	<i>9 vs. 10</i>	<i>10 vs. 11</i>	<i>11 vs. 12</i>	<i>12 vs. 13</i>	<i>13 vs. 14</i>
1	0.018	0.031	0.004	0.054	0.023	-0.017	-0.090	0.004	0.022	0.026	0.273	-0.307	-0.087	0.236
2	-0.075	0.006	-0.043	0.132	-0.047	0.010	0.016	-0.051	0.043	-0.026	0.354	-0.365	0.099	-0.054
3	-0.026	0.025	0.006	-0.002	-0.053	0.067	-0.151	0.062	0.119	-0.031	0.089	-0.092	0.032	0.068
4	-0.004	0.004	0.089	0.008	0.049	0.052	-0.100	-0.089	-0.154	0.341	0.354	-0.522	0.214	-0.149
5	-0.044	-0.004	-0.050	-0.026	-0.108	0.062	0.116	-0.125	-0.055	-0.144	0.315	-0.161	0.136	-0.070
6	0.018	0.022	-0.021	0.070	0.069	-0.036	0.162	-0.232	0.167	-0.140	0.365	-0.272	-0.034	0.021
7	0.004	-0.126	0.128	-0.123	-0.006	0.007	-0.026	0.061	0.027	0.080	0.094	-0.318	0.235	-0.081
8	-0.093	0.036	-0.075	0.025	-0.045	0.037	0.004	0.044	-0.092	0.084	0.177	-0.239	-0.016	0.136
9	-0.030	0.085	-0.035	-0.026	0.114	-0.112	-0.096	0.147	-0.070	0.062	-0.002	-0.107	0.235	0.042
10	0.010	-0.048	0.079	0.009	0.089	-0.022	0.155	-0.239	0.032	0.038	0.336	-0.417	0.127	-0.009
11	-0.042	-0.054	0.018	0.022	0.043	0.031	0.108	-0.173	0.115	-0.115	0.151	-0.109	0.007	-0.050
12	0.051	-0.031	-0.001	0.104	0.012	0.018	-0.151	0.005	0.111	-0.080	0.276	-0.208	0.074	0.016
13	-0.008	-0.079	0.104	-0.099	0.115	-0.094	0.120	-0.008	0.003	-0.052	0.063	-0.052	-0.104	-0.049
14	-0.052	-0.043	0.106	-0.097	0.100	-0.080	-0.090	-0.124	0.103	0.193	0.162	-0.346	0.056	-0.176
15	0.034	-0.061	0.038	0.008	-0.046	-0.009	-0.088	0.112	-0.020	0.048	0.171	-0.194	0.069	0.243
16	-0.072	-0.050	-0.062	0.033	-0.164	0.147	-0.224	0.054	-0.090	0.214	-0.097	-0.113	0.057	0.141
17	0.028	-0.039	0.077	-0.027	0.116	-0.032	0.115	-0.059	-0.104	0.111	0.144	-0.155	0.018	0.020
18	0.055	-0.064	0.080	-0.048	0.081	-0.063	0.082	0.001	0.101	-0.039	0.325	-0.383	-0.047	0.034
19	-0.025	0.000	0.018	0.014	-0.035	0.006	0.173	-0.152	-0.016	0.007	0.227	-0.240	-0.016	0.043
20	0.015	0.032	0.001	0.046	0.015	-0.015	-0.052	-0.124	0.354	-0.151	0.363	-0.391	0.177	-0.108
21	-0.051	-0.052	-0.015	0.046	-0.074	-0.055	-0.126	0.108	0.039	0.100	0.091	-0.234	0.055	0.272
22	0.054	0.041	0.014	0.029	0.013	-0.027	0.091	-0.130	0.184	0.013	0.184	-0.424	0.112	-0.007
23	-0.016	-0.001	-0.052	-0.006	-0.109	0.018	0.140	-0.324	0.190	-0.081	0.056	-0.074	0.011	0.131
24	0.075	0.020	0.036	-0.098	0.164	-0.044	0.095	-0.165	0.196	-0.184	0.409	-0.478	0.159	-0.099
25	0.009	0.021	0.016	0.033	0.025	-0.047	0.163	-0.208	0.028	0.104	0.110	-0.062	-0.077	0.056
26	0.035	0.007	0.041	-0.045	0.053	-0.049	0.083	-0.076	0.074	0.096	0.21	-0.411	-0.006	0.092

p-value cut-offs reflect test-wise alpha level adjustments using Bonferroni corrections (α / c), given 364 tests: * $p < 0.000137$ ($0.05 / 364$), ** $p < 0.000027$ ($0.01 / 364$),

*** $p < 0.0000027$ ($0.001 / 364$).

[†] Item number refers to the corresponding item in Table 9.

Values reported are Cohen's d and stars (*) indicate that the difference in item mean responses is significant according to an independent samples t-test. No tests reached significance.

Table 9*Structural Invariance Test Results of First-Order Structural Model for No Reminder vs. One or More Reminders*

	χ^2	<i>df</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>	<i>SRMR</i>	<i>AIC</i>	<i>BIC</i>	$\Delta\chi^2$ [†]	Δdf [†]
Model 1: Constrained Measurement Parameters	47655.67***	9118	0.918	0.915	0.038	0.069	928998.6	934211.1	-	-
Model 2: Constrained Latent Variances	47782.57***	9143	0.918	0.915	0.038	0.070	929075.5	934120.9	126.90***	25
Model 3: Constrained Latent Covariances	48202.39***	9315	0.917	0.916	0.038	0.070	929151.3	933047.3	419.82***	172
Model 4: Constrained Regression Paths	48294.19***	9361	0.917	0.916	0.038	0.071	929151.1	932739.7	91.80***	46

N = 5900; *** *p* < 0.001; † Statistics compare model fit to the prior nested model.**Table 10***Structural Invariance Test Results of Second-Order Structural Model for No Reminder vs. One or More Reminders*

	χ^2	<i>df</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>	<i>SRMR</i>	<i>AIC</i>	<i>BIC</i>	$\Delta\chi^2$ [†]	Δdf [†]
Model 1: Constrained Measurement Parameters	70344.58***	9509	0.871	0.871	0.047	0.087	950905.5	953505.1	-	-
Model 2: Constrained Latent Variances	70640.97***	9539	0.870	0.870	0.047	0.087	951141.9	953541.0	296.38***	30
Model 3: Constrained Latent Covariances	70647.93***	9540	0.870	0.870	0.047	0.087	951146.8	953539.3	6.96**	1
Model 4: Constrained Regression Paths	70678.93***	9544	0.870	0.870	0.047	0.088	951169.8	953535.5	31.00***	4

N = 5900; ** *p* < 0.01, *** *p* < 0.001; † Statistics compare model fit to the prior nested model.

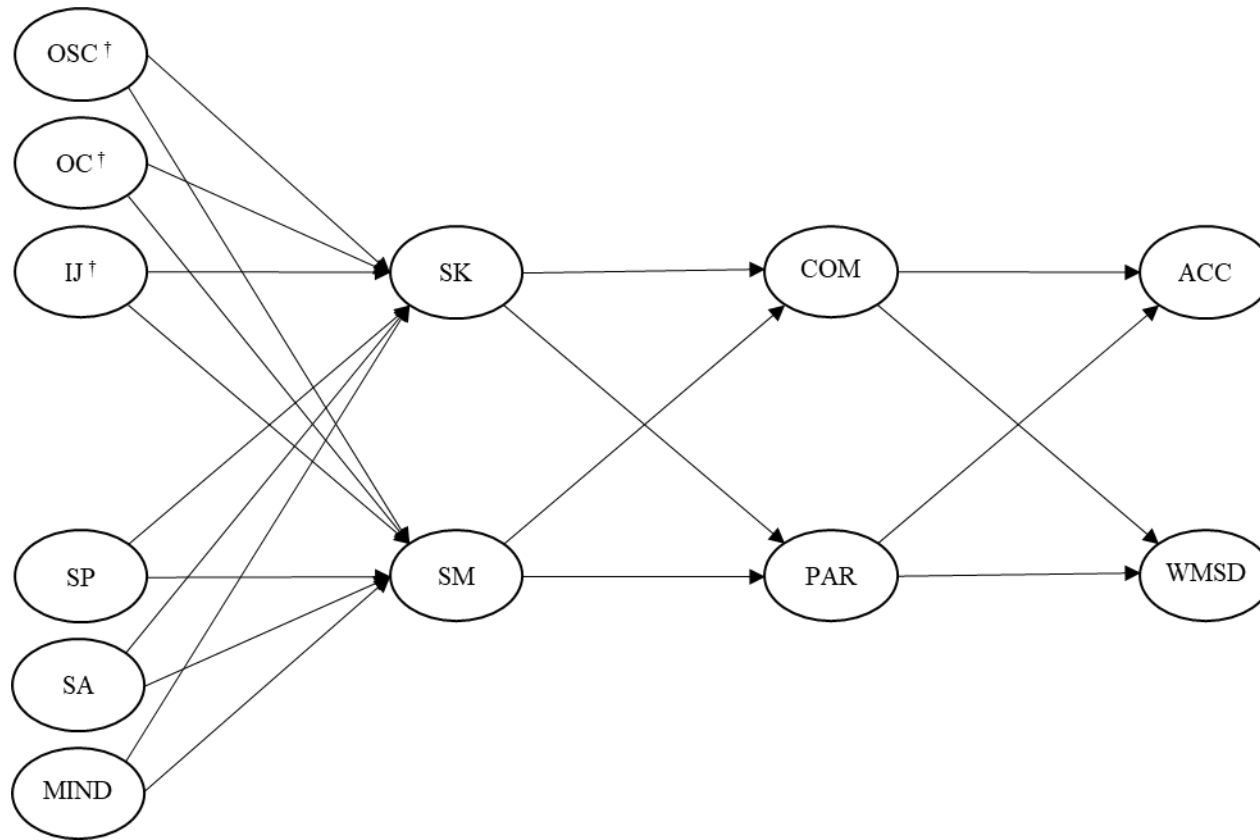


Figure 1. *First-Order Model of the Effects of Distal Situation Factors, Distal Person Factors, and Proximal Person Factors on Safety Behavior and Outcomes.*

Note 1: Direct effects between mediated constructs (e.g. between OSC and COM) are not included in this representation of the tested model for visual simplicity. Such effects were included in the analyses.

Note 2: Survey items are not included in this representation of the tested model for visual simplicity.

Note 3: OSC = organizational safety climate; OC = open communication; IJ = informational justice; SP = safety priority; SA = safety attitude; MIND = mindful attention; SK = safety knowledge; SM = safety motivation; COM = safety compliance; PAR = safety participation; ACC = accidents; WMSD = work-related musculoskeletal disorders.

†All other constructs in the survey not displayed in this representation were included with this group of constructs in the tested model. These were not included in this representation for visual simplicity.

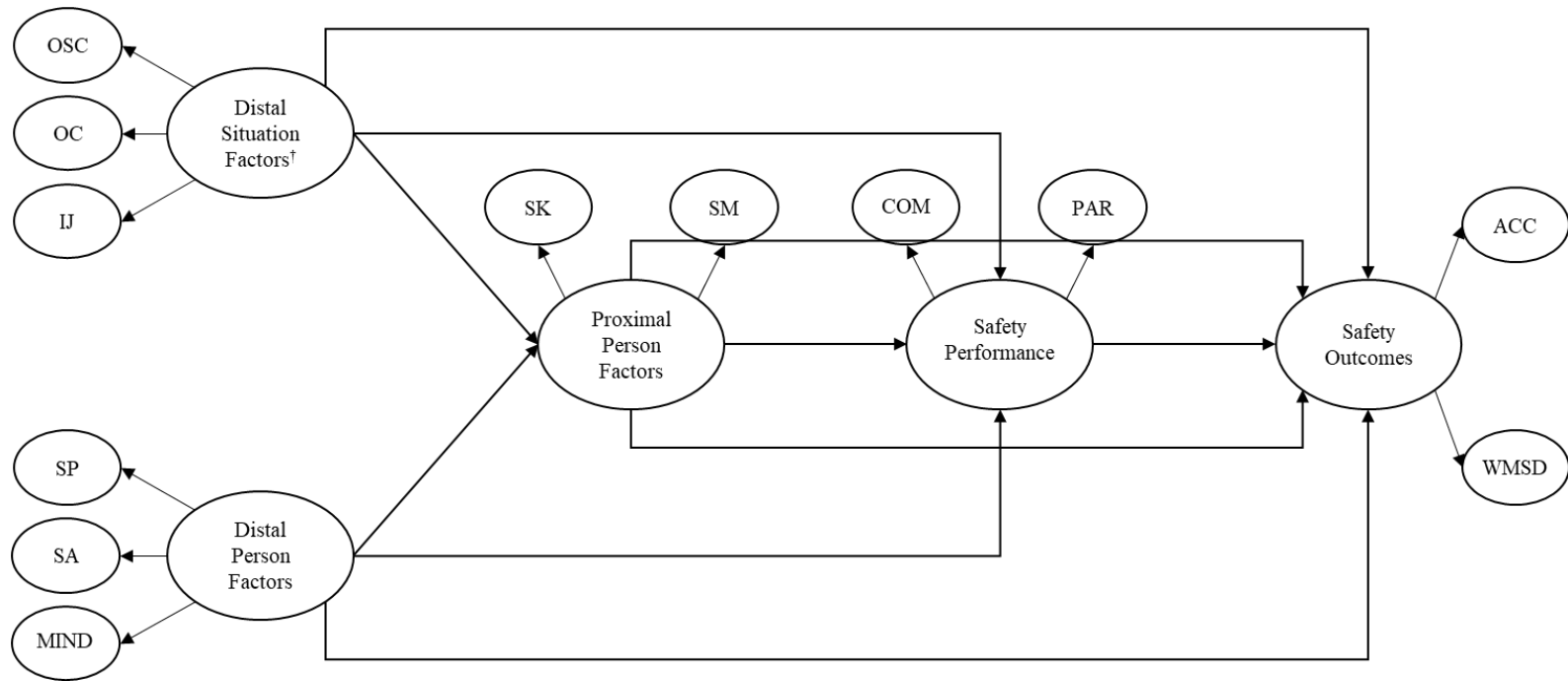


Figure 2. *Second- Order Model of the Effects of Distal Situation Factors, Distal Person Factors, and Proximal Person Factors on Safety Behavior and Outcomes.*

Note 1: Survey items are not included in this representation of the tested model for visual simplicity.

Note 2: OSC = organizational safety climate; OC = open communication; IJ = informational justice; SP = safety priority; SA = safety attitude; MIND = mindful attention; SK = safety knowledge; SM = safety motivation; COM = safety compliance; PAR = safety participation; ACC = accidents; WMSD = work-related musculoskeletal disorders.

†All other constructs in the survey not displayed in this representation were loaded onto this second-order construct in the tested model. These were not included in this representation for visual simplicity.