International Journal of Aviation, Aeronautics, and Aerospace

Volume 1 | Issue 4 Article 1

10-2-2014

Measurable Outcomes of Safety Culture in Aviation - A Meta-Analytic Review

Benjamin J. Goodheart
AirSure Limited/Embry-Riddle Aeronautical University Worldwide, goodheb@erau.edu

MaryJo O. Smith
Ypsilon Associates/Embry-Riddle Aeronautical University Worldwide, drmaryjosmith@gmail.com

Measurable Outcomes of Safety Culture in Aviation - A Meta-Analytic Review

Cover Page Footnote
Corresponding author: Benjamin Goodheart, bgoodheart@airsure.com

Measurable Outcomes of Safety Culture in Aviation: A Meta-Analytic Review

Though the concept of safety culture has existed in some iteration or another for over 30 years, it has with relative rarity been the subject of empirical studies seeking to to examine the efficacy of a positive safety culture with respect to measurable safety performance. Given that civilian aviation has long been at the fore of the study and development of human-centric safety programs, this scarcity of data and reported effect size is somewhat surprising (O'Connor, O'Dea, Kennedy, & Buttrey, 2011). This study addressed this problem and contributed to the advancement of the field of aviation safety as it sought to consolidate previous efforts into a meaningful, scientifically-based consensus. In addition, this study identified and quantified the effect size associated with different levels and different measures of safety culture with respect to resultant safety performance, with empirically-derived weighting of studies to support conclusions. This research provides a foundation upon which further inquiry into safety culture interventive policies and procedures can be based and lends validity to those instruments that contribute most to the understanding of organizational safety culture as it relates to measured safety behavior.

Problem Statement, Research Question, and Hypotheses

This study addressed the research question: does a positive safety culture in an aviation setting, as measured by some recognized instrument, demonstrate significant relationships with safety performance? Safety performance was contextually defined as those behaviors and outcomes that are generally recognized indicators of the level of safety of an organization and included but was not limited to incident and accident reports, lost time, reported injuries, annual cost of incidents/accidents, accident/incident rate, audit performance, timeliness of hazard report resolution, and others. A priori, we hypothesized that safety culture would demonstrate a negative correlation with reported incidents and accidents (positive measures of safety culture/climate related to lower rates of incidents and accidents). This same hypothesis was extended to include leading indicators of safety, though the correlation here was hypothesized to be positive (higher measure of positive culture correlates to positive safety performance). The purpose of this research was threefold: to perform a systematic review of relevant literature, to identify via meta-analysis the effect of safety culture on safety performance in an aviation operational setting, and to make recommendations for further research and improvements to validity or reliability of safety culturel measurement instruments.

Review of Literature

The phrase *safety culture* became part of the common safety lexicon following its use in the International Atomic Energy Agency's report on the Chernobyl disaster of 1986 (Makino, 2006). Though several definitions of safety culture have been put forward in an effort to clarify or constrain the concept, the definition continues to evolve alongside the safety industry. Perhaps the most seminal of descriptions of the concept of an organizational safety culture is that put forth by the Advisory Committee on Safety in Nuclear Installations and suggested by Cox and Flin (1998) as the most widely used of the many definitions in the literature: "the safety culture of an organization is the product of individual and group values, attitudes, perceptions, competencies, and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization's health and safety management" (Health and Safety Commission, 1993, p. 23). While this definition suffices as a means of narrowing focus within the context of this study, it is certainly not a consensus nor is it the only valid definition of safety culture.

Alongside the literature that addresses safety culture is the substantially similar concept of climate. Organizational climate, and safety climate by extension, tends to be the more prevalent measure in questionnaires and surveys that seek to measure current perceptions and attitudes with respect to safety. Illustrating its similarity to culture is Zohar's (1980) seminal definition of climate: "…a summary of molar perceptions that employees share about their work environments… a frame of reference for guiding appropriate and adaptive task behaviours (sic)" (p. 96). As in the literature specific to safety culture, no consensus view of the appropriateness of measurement of safety climate exists as a reliable predictor of safety performance, especially in the context of an aviation organization.

Cox and Flin (1998) likened the relationship between the constructs of safety culture and safety climate to measurement of personality (culture) and mood (climate). Debate continues over the semantic and operational differences between culture and climate as used to describe certain organizational characteristics within an operational safety framework. However, in the context of this research the terms were used interchangeably with the intended meaning being rooted in the International Atomic Energy Agency's vision of culture (Health and Safety Commission, 1993) while being more temporally aligned with the foregoing description of climate, and with the understanding that the two ideas are conceptually complimentary. This broad, encompassing view of safety culture has been supported empirically (Hoffman & Mark, 2006) in the context of relating safety performance to sub-dimensions of safety and functions in the present application to retain an inclusive definition for the purpose of deriving a population for meta-analysis.

Safety performance remains subject whose dimensionality has not been widely explored. However, it is an important concept to specify considering that studying safety climate, an equally variable construct, as an antecedent to safety performance has been the subject of research over the past three decades and forms the basis for the present study. The concept of performance encompasses the actual behaviors performed by individuals at work. Although accident rate in many aspects of the aviation industry is too low to offer a sensitive measure of safety performance, other leading indicators of performance collected by aviation operators may supplement incident rate as a measure of safety intervention effectiveness (O'Connor, et al., 2011). Burke, Sarpy, Tesluk, and Smith-Crowe (2002) identified four safety performance factors through confirmatory factor analysis: using personal protective equipment, engaging in work practices to reduce risk, communicating health and safety information, and exercising employee rights and responsibilities. Though the Burke et al. (2002) research provides a foundation for defining safety performance, its conception is arguably limited in generalizability to other industries, despite the importance of the identified relationship of knowledge and training to each of these factors. Safety knowledge, participation, and compliance were identified as constructs of performance by Griffin and Neal (2000); and Hoffman and Stetzer (1996) considered similar group-level and individual behaviors. In the present study, a holistic view of performance indicators was adopted in order to more inclusively capture the relevant studies.

Though safety culture (or climate) was the primary focus of this research, it is not the only probable antecedent to safety performance. Perhaps the most pronounced, and logical, predictor of safety performance is the set of hazards associated with the operational environment. We attempted to control for these by limiting the scope of analysis to studies within the aviation operations sector, thus assuring at least a baseline homogeneity in the scope of hazards under study. Proactive safety strategies, such as safety-related training or policies and procedures, are also a likely predictor of lower accident rates and positive safety behaviors (Hayes, Perander, Smecko, & Trask, 1998). For the purpose of this research, only safety culture and climate were of interest as antecedents, though safety commitment is arguably an antecedent of culture itself (e.g., Clarke, 1999; Human Engineering, 2005), and some conclusions may be drawn about the inherent link to the present research. Figure 1 illustrates this conceptual link and the focus within this research on only the relationship between culture and performance.

Figure 1. Relationship between antecedents of safety culture and performance.

Meta-Analysis Criteria

Over the past three decades, researchers have struggled to identify clear evidence that links safety climate to safety performance. Cooper and Phillips (2004) identified four research directions within the field of safety culture: the design of psychometric measurement instruments, exploration of links between safety culture and organizational culture, identification of theoretical antecedents of safety behavior, and the study of the relationship between safety climate and actual safety performance. Within the latter segment, researchers have studied the link between safety climate and behavior in chemical processing (Hoffman & Stetzer, 1996), manufacturing (Zohar, 2000), nuclear waste handling (Smith-Crowe, Burke, & Landis, 2003), construction (Glendon & Litherland, 2000), and offshore oil and gas (Mearns, Whitaker, & Flin, 2003), among others. Within this already small body of research, only a few studies were found to have addressed aviation specifically. These studies are discussed in turn, with specific attention paid to sources of bias, sample methods, reliability, validity of the measurement instrument, effect size, and direction of the relationship between the concepts of culture, climate, or attitudes and safety performance. Specific inclusion criteria that dictated selection of the studies shown in Table 1 are addressed in the sections that follow. Of note is that few studies in the survey of literature contained the effect size information necessary for inclusion in a meta-analysis.

Table 1

Studies Initially Selected for Inclusion in Meta-Analysis

Authors (Data)	Scale Used	Outcome Measure	Sample	N
Brittingham (2006)	Maintenance Climate Assessment Survey	None used	US Navy maintainers	126,058
Helmreich, et al. (1986)	Cockpit Management Attitudes Questionnaire	Check pilot-rated performance	Pilots from three major US airlines	114
Hernandez (2001)	Maintenance Climate Assessment Survey	Incident rate	US Navy maintainers	2,180
O'Connor, et al. (2011)	Command Safety Assessment Survey	Mishap rate	US Navy aircrews	23,442
Sexton & Klinect (2001)	Flight Management Attitudes Short Survey	Errors (undesired aircraft state), error chains, policy violations, crew effectiveness measured by LOSA observers	Major US airline crews	163

Studies Included for Meta-Analysis

Brittingham's (2006) research sought to address the relationship between US Naval aviation mishaps and squadron maintenance safety climate. The Maintenance Climate Assessment Survey (MCAS) data collected between August 2000 and August 2005 were used to investigate whether MCAS results were predictive of mishap likelihood for Naval and Marine Corps Aviation squadrons. The validity of the MCAS and its safety categories was examined in this study through principal component analysis and principal axis factoring as well as through ANOVA and scale discrimination and correlation. In this phase of the research, the study author revealed that the MCAS appeared to load onto a single factor that explained roughly 50 percent of the variance. Two other factors showed only minimal influence, contrary to Baker's (1991) research that developed the original MCAS, and in any event, Brittingham (2006) was unable to show that the MCAS was a valid tool for evaluating the Model of Organizational Safety Effectiveness it was purported to measure. Consequently, the intended analysis of MCAS results' relationship to mishaps was not undertaken, and conclusions of the research focused primarily on recommendations for survey review. This conclusion was puzzling given the widespread use of the MCAS, though no definite indication was given as to whether the MCAS instrument had undergone structural changes that may have affected its utility as an indicator of safety culture consistent with the factors of interest.

Helmreich, Foushee, Benson, and Russini (1986) measured attitudes of flight crew members against flying performance evaluations. Culture and climate were not explicitly mentioned as elements of measurement; however, attitudes were a principal component of culture and climate (Diaz & Cabrera, 1997). Because the instrument used in the study (Cockpit Management Attitudes Questionnaire [CMAQ]) was based on Likert scale responses, the potential for central tendency bias was present, though it was not addressed by the study authors. The CMAQ has been shown to be both reliable and sensitive, and to have good content and predictive validity (Gregorich, Helmreich, & Wilhelm, 1990). Study subjects were selected from the existing CMAQ database of 658 respondents. From this group, 114 pilots currently flying Boeing 727 and 737 aircraft were chosen for evaluation of performance. Performance assessment was completed by check pilot raters who observed line flight operations involving subject crew and rated performance on a Likert-scale questionnaire. Discriminant analysis was used to compare subjects who were rated as superior with those rated as below average. The reported Wilk's lambda was 0.36, and the associated chi-square of 36.78, $df = 18$, $p = 0.006$, corresponds to the significant predictive power of the measured attitudes toward cockpit safety and operational performance.

Hernandez (2001) analyzed results from the MCAS taken from US Naval aviation maintenance facilities and submitted via the internet. The use of internet-reported survey results introduces an element of convenience to the sampling method, and the concurrent availability of the instrument in paper and electronic forms could have introduced bias as only the electronic records were used in the study. The researcher did note that no significant differences appeared to exist between the paper-and-pencil version of the survey instrument and the electronic one. Despite the reservations with respect to validity raised by Brittingham (2006), Hernandez referred to recent revalidations of the MCAS by two researchers in independent studies only a year prior to his research. Results from the MCAS were statistically evaluated through ANOVA and MANOVA as appropriate against the components of the Model of Organizational Safety Effectiveness (MOSE) used by the US Navy. A total of 1,731 surveys were selected from the total of 2,180 and were used for analysis. A number of analyses were conducted, and as a measure of the relationship between MCAS score and unit-level incidents, linear regression was conducted with unit mean MCAS score as the independent variable and incident rate as the dependent variable. For all models, the estimated slope was negative (-61.38, $t = -0.36$) and indistinguishable from random effects ($p = 0.720$), $R^2 = 0.005$, F = .131 ($df = 1$). The study author noted that sample size may have affected this relationship as well as the theorized rise in safety awareness following an incident or accident.

O'Connor, Buttrey, O'Dea, & Kennedy (2011) used logistic regression modeling to investigate whether results of the Command Safety Assessment Survey were useful in differentiating US Navy squadrons with recent mishaps and those without. Survey responses from 23,442 US Naval aircrew members were used in the regression models tested in this research. The study authors carefully discuss discriminate and construct validity of the measurement instrument. Similar discussions of discriminate and construct validity were noticeably lacking from much of the reviewed literature on safety culture or climate, and this discussion substantially addressed primary threats of bias in the research. To correct for central tendency bias that may have been present where many respondents gave the very same numeric response to almost all survey items, the researchers replaced the responses with the difference between that response and the mode for that respondent. Potential bias in the classification of mishap severity was addressed by the use of the Naval Safety Center summaries and severity ratings coupled with independent review by two researchers. Regression models were sequentially evaluated using the Akaike Information Criterion value to determine what term was added. Sequential regression methods may not always result in a sensible model, but the study authors noted, “our goal is not to select the “correct” set of variables, nor to obtain accurate estimates of individual regression coefficients” (O'Connor et al., 2011, p. 16). Instead, the study sought primarily to identify a *simple* model with reasonable predictions. The final of ten logistic regression model iterations resulted in $R^2 = 0.206$, $df = 4$, 0.831 area under the curve, and a Hosmer-Lemeshow p-value of 0.98. This model included only factor one (personnel leadership; items 4, 9, 13, 16, 17, 19, and 48 – items 4, 13, and 17 were shown in the regression results to be the only three items predictive of mishaps) results and omitted the two communities that experienced no mishaps at all during the period under investigation.

Sexton and Klinect (2001) explored the relationship between pilot performance and self-reported perceptions of organizational culture among airline flight crews. They echoed a recurring theme from the literature in noting that the link between organizational culture and flight or safety performance was largely anecdotal. Data were collected during Line Operations Safety Audits (LOSA) on randomly-selected revenue flights at a major airline using the Flight Management Attitudes Short Survey (FMASS), the result of an effort to shorten the Flight Management Attitudes Questionnaire, for which validity and reliability was established in research by Helmreich, Merritt, Sherman, Gregorich, and Wiener (1993) and by Helmreich and Merritt (1998). As in other studies addressed here, the survey instrument used a Likert scale and may have introduced bias via central tendency in responses. Bias was controlled in LOSA observer responses through training and calibration to a group norm of .80 or higher. An omnibus test for safety culture yielded a significant Hotelling's Trace F (12, 124) = 2.39, $p \leq 0.008$. For job attitudes, the omnibus test was marginally significant, Hotelling's Trace $F(12, 84) = 1.74$, $p \leq 0.074$. The use of two separate MANOVAs contributed to an inflated alpha, and the MANOVAs did not include nearly a third of the flight crews (those that did not commit errors recorded in the LOSA observation).

Methods and Procedures

A review of literature was conducted by searching commercial databases including Science Direct, ProQuest, ProQuest Dissertations and Theses, Google Scholar, and PsychInfo for research articles, doctoral dissertations, conference proceedings, and manuscripts addressing the role of safety culture as an antecedent to safety performance, primarily but not exclusively in an aviation operational setting. Keywords included the terms *aviation, safety, culture, climate, outcome, performance, injury, accident, and incident* in Boolean logical combinations to identify relevant articles, specifically those that were published within the last ten years. This timeframe was loosely imposed, but was intended to structure the search given the emergence of safety management systems in aviation in that time period and the accompanying increase in awareness of the concept of safety culture. A manual search of reference lists of included articles was also used to supplement the electronic database search. These methods resulted in the identification of approximately 158 articles, including those unrelated to aviation, from which the meta-analysis sample was selected.

Inclusion Criteria

Studies were selected for further analysis in accordance with The Cochrane Collaboration's guidelines for conducting a systematic review of research literature. The Cochrane Collaboration is an international not-for-profit organization focused on the furtherance of evidence-based healthcare that publishes guidance concerning systematic reviews and meta-analytic studies (The Cochrane Collaboration, n.d.). While the Cochrane review processes generally relate to human healthcare, the process established a scientific rigor and transparency to the methods used for selection of studies for review, thus reducing bias. This reduction in bias and structure, which is not specific to any one field of research, led the authors to use to apply the Cochrane process to aviation for what appears to be the first time.

To identify content appropriate for inclusion in the meta-analysis, abstracts for each article were reviewed. Those studies that did not identify data because of a focus on theoretical investigation or literature reviews were excluded from consideration for further analysis. This initial qualification of articles for study reduced the population to 51 studies.

Only those studies that deployed a safety culture or climate measurement instrument to evaluate these measures in participants within aviation organizations were included. Furthermore, studies outside the scope of flight operations, aircraft maintenance, or aircraft ground handling and service were also excluded. Review of the remaining studies by both authors excluded those that did not address at least one of the previously discussed safety performance constructs in conjunction with safety culture or climate. Next, studies that utilized non-unique samples were excluded when it was determined that no unique relationships were presented in a subsequent analysis. In these cases, only the primary work was included for further inquiry. Finally, and most importantly, in cases where reported results were insufficient for the calculation of an effect size, the study was excluded. Application of these criteria resulted in five studies for possible inclusion in the meta-analysis. Only four studies were truly appropriate for inclusion upon further review, as Brittingham (2006) elected not to investigate the link between culture measurement and safety performance on the basis that the instrument was found to be flawed. The meta-analysis was subsequently accomplished with two separate omnibus measures from Sexton and Klinect (2001) that represented safety culture and job attitudes as separate constructs. The literature review provided sufficient support for the interconnectedness of these constructs with respect to safety climate and culture to validate the inclusion of the job attitudes measure, the results of which are addressed in the following discussions.

Data Coding

Coding procedures were not as extensive as in some examples where experimental designs are used or several moderating variables are present. In this case, the authors reviewed the studies as outlined in the preceding section, and the very small number of studies that met the basic inclusion criteria made creation of a coding manual unnecessary. The authors reviewed the list of potential studies for analysis independently, and any disagreement was resolved through discussion. The lack of reported or calculable effect size measurements in the bulk of studies resulted in identical lists from both authors, and discussion was generally limited to inclusion of both measures as reported in Sexton and Klinect (2001) and to what extent inclusion may violate assumptions of independence. Similarly, the final list of studies was reviewed to code, or extract information including at least:

- APA style reference (all studies),
- measurement instrument type,
- aviation context (all studies),
- study design,
- subjects,
- *N*,
- effect size and calculation method, and
- eligibility for inclusion.

Effect Sizes

In contrast to a narrative review, many of which focus on or base conclusions largely on *p*-values, this research focused instead on effect sizes as a function of size and direction of the relationship between safety culture and safety performance. Whereas *p*-value only indicates *statistical* significance and that the effect size is probably not zero, it does not necessarily serve as an indicator of the size of the effect. As discussed previously however, the reported results of the studies selected for inclusion were heterogeneous. To obtain a homogenous measure of effect size, reported results were converted to *r* as a measure of correlation. Computation of *r* was completed as follows:

- Helmreich, Foushee, Benson, and Russini (1986) reported Wilk's lambda of 0.360. As a multivariate measure of correlation, Wilk's lambda can be converted to a canonical correlation (analogous to *r*) using Equation 1.

$$\sqrt{1-\lambda} \tag{1}$$

Using this computation, the correlation associated with Helmreich, et al.'s (1986) results was calculated as 0.800.

- Hernandez (2001) reported results of the linear regression as $R^2 = 0.005217$, resulting in a computed R of 0.072.
- O'Connor, et al. (2011) reported the logistic regression result $R^2 = 0.206$, giving a computed R of 0.454.
- Sexton and Klinect (2001) reported an omnibus result as Hotelling's trace with $F(12, 124) = 2.39$. Critical F-value was computed as 1.83, which indicated that the result was significant as reported in the article. Conversion to r was accomplished as in Equation 2:

$$\sqrt{\frac{\mathrm{df_n}F}{\mathrm{df_n F+df_d}}} \tag{2}$$

The computed r-value using Equation 2 was 0.433 (included in Figures 2 and 3) for safety culture and 0.446 for job attitudes (included in Figure 3 only).

Results

Because the variance depends strongly on the correlation, the meta-analysis did not perform syntheses on the correlation coefficient itself, but rather on the Fisher's Z as computed using the *Comprehensive Meta-Analysis (CMA)* software package (Borenstein, Hedges, Higgins, & Rothstein, 2009). Figures 2 and 3 illustrate the forest plots for the random effects model with and without the *job attitudes* result from Sexton and Klinect (2001). The random-effects model

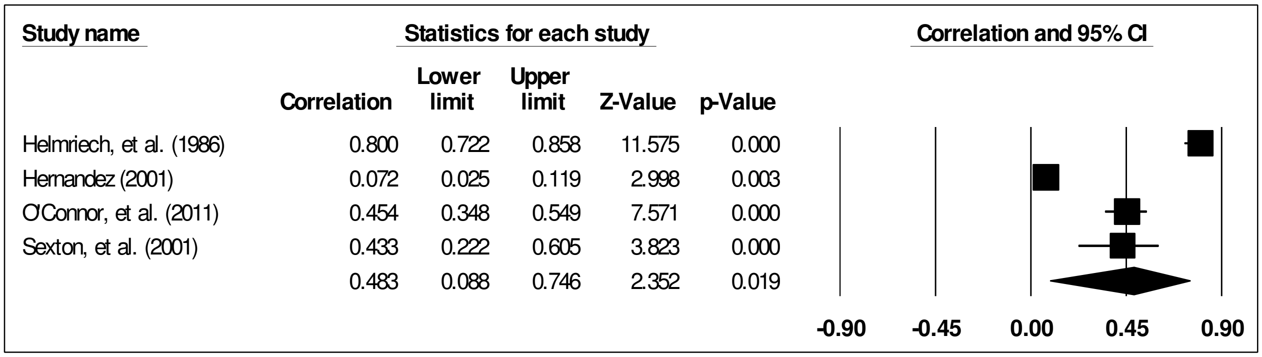

Figure 2. Random effects meta-analysis schematic (without Sexton, et al. (2001) attitudes).

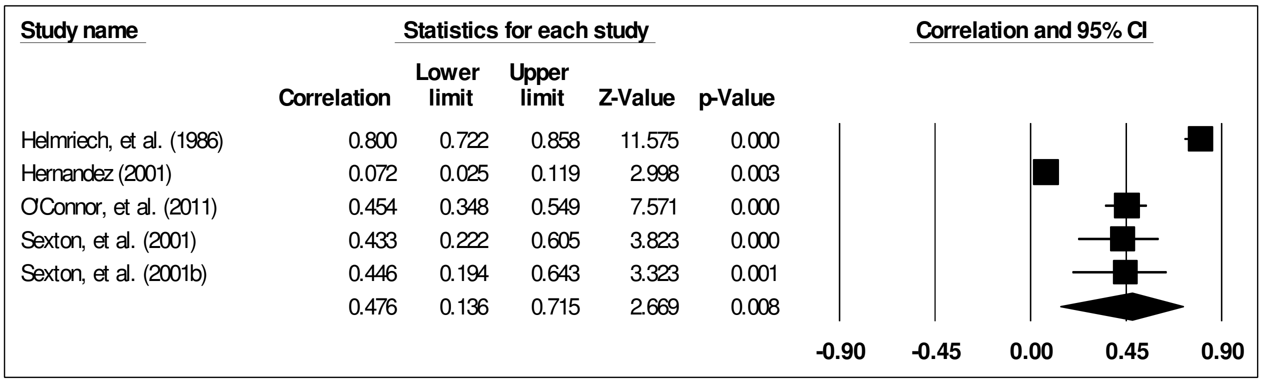

Figure 3. Random effects meta-analysis schematic (with Sexton, et al. (2001) attitudes).

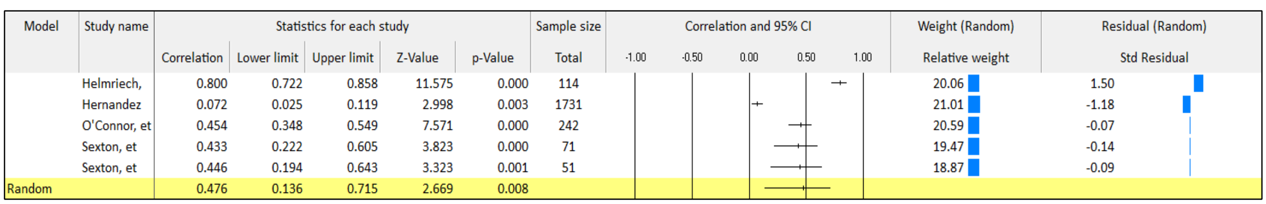

Figure 4. Random effects meta-analysis results (with Sexton, et al. (2001) attitudes).

Table 2

Random Effects Model Measures of Heterogeneity

Heterogeneity			
Q-value	df (Q)	P-value	I-squared
146.237	4.000	0.000	97.265

was used, as the fixed-effects assumption of homogeneity of true effect size across studies was untenable. Differences in study participants, as well as the underlying construct definitions, contributed to differing effect sizes characterizing each study, though inclusion criteria ensured that the effect sizes were at least similar.

Figure 4 provides sample size, native and computed statistics for each study, correlation and confidence intervals, and study weights under the random effects model. The random effects weights shown here have a much narrower range than the fixed effects counterpart because the supposition that all studies share the same true effect size was rejected in favor of assuming that each study may have had a different true effect. Table 2 provides several measures for identifying and quantifying the variation in true effect sizes among studies.

Discussion

The meta-analytically calculated summary effect size using the random effects model and including both safety culture and job attitudes from Sexton and Klinect (2001) was 0.476, $p = 0.008$. This result is a measure of correlation and was interpreted here as a Pearson's coefficient. Thus, the summary effect size showed a statistically significant moderate positive effect. This result was substantially similar to the one obtained without the addition of job attitudes. Though inclusion of attitudes as a separate study may compromise independence, the results do not appear to be affected, with a correlation coefficient of 0.483, $p = .019$ as the alternative result.

Some discussion of heterogeneity is warranted given the large value of Q in Table 2. Q is a statistic that represents the ratio of observed variation to within-study error – a measure of dispersion (Borenstein, et al., 2009). In this context, Q represented a first step in determining homogeneity across studies in the analysis. Because $p < .001$, it can be inferred that the true effects did indeed vary, though it is not a direct measure of the actual amount of dispersion given that the results were sensitive to the number of studies included in the analysis. The reported value of *Q* indicated that excess dispersion in the study was not zero, as *Q* was larger than *df*. The large reported I^2 meant that further investigation of the reasons for the variance was supported. According to general guidelines presented by Higgins, et al. (2003), the reported I^2 value was rather high, which indicated most of the observed variance was real. Estimated tau and tau-square (T and T^2) remove the dependence on the number of studies from the *Q* estimate so that the variance and SD of the true effect could be estimated. The positive T^2 (0.178) was expected given the value of *Q* and is an indication of the absolute amount of variance as opposed to a ratio. The computed value of *T* of 0.422 described the distribution of effect sizes around the mean effect, and in this case, the dispersion was characterized as rather wide. This conclusion was based on *T* as an estimate of standard deviation, and the large value of *T* computed here meant that most effect sizes fell within a range that extended beyond the possible values of the mean correlation.

Whether or not the effect size could be considered reliable was another subject altogether, and the small number of included studies was cause for concern (Ellis, 2010). Though meta-analyses can generally be expected to generate higher statistical power and minimize bias, this naturally depends on the availability and quality of the included studies. In the present case, significant evidence existed to support a positive link between safety culture and safety behavior. However, unpublished studies likely existed that were not available for inclusion, and evidence existed to support availability bias as having had a bloating influence on effect size. When graphed, the dispersion of the included studies should describe an inverted funnel shape (Ellis, 2010). In this case, both the number of studies and the wide dispersion made it difficult to combat the effects of availability bias and over-inflated effect size as a result (see Figure 5 for the random effects funnel plot). Computation of fail-safe *N*, the number of studies with

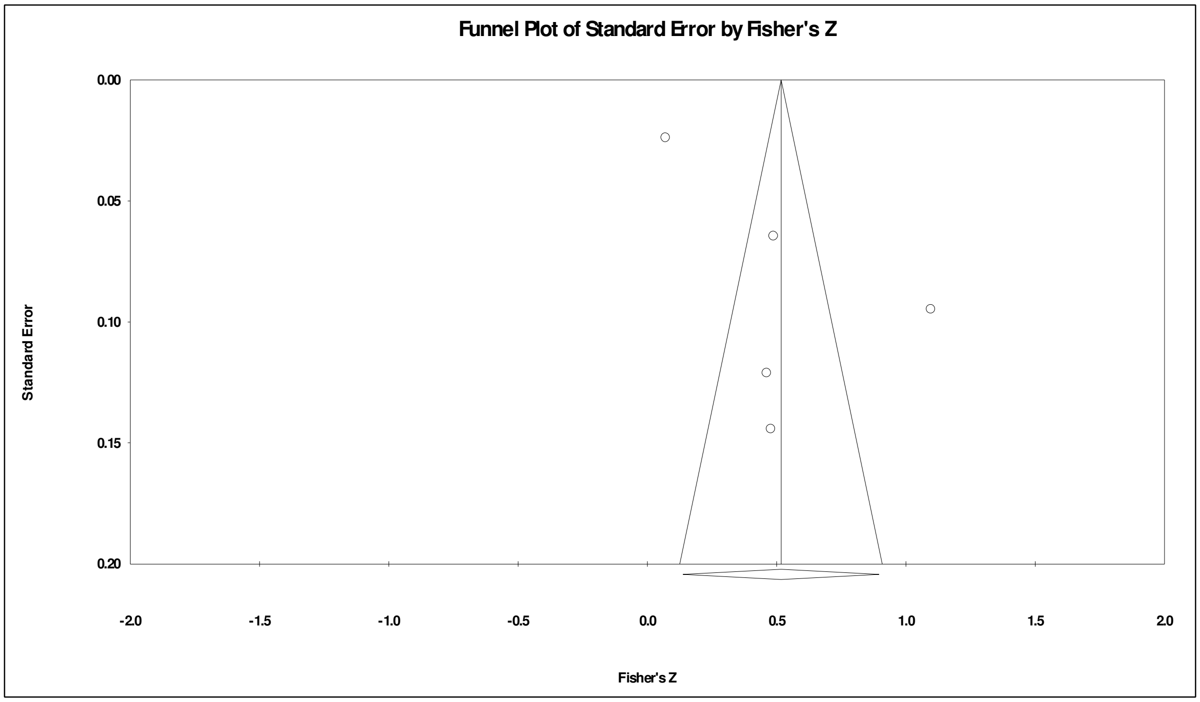

Figure 5. Funnel plot of included studies.

conflicting evidence required to overturn the summary effect, indicated that only five additional, contradictory studies (5.443) would nullify the conclusion reached here. In addition to falling far short of the minimum recommended fail-safe N ($5k + 10$), this result was rather revealing of the intolerance of the outcomes here to null findings. The small number of studies, large degree of dispersion, and tenuous summary effect with respect to confounding results combined to restrict interpretation of these results to simple observations rather than broader empirical conclusions.

Limitations

The results of the meta-analysis must be weighed against the small size of the sample as well as the methodology involved. Although our methods arguably may have allowed for higher statistical power, the number of studies in the existing literature remains so few that the meta-analytically derived effect size was potentially unstable. From a practical perspective, the scarcity of data points to a need for further study of the relationship between safety culture, climate, and performance, as well as further investigation of the directionality of mediating relationships. The authors look to Helmreich's (and others') work on the Cockpit Management Attitudes Questionnaire (CMAQ) as one solution to this problem. In the CMAQ example, a measurement instrument was made available through contact with the study authors in exchange for the resultant data. In the present case, the adoption of a unified instrument with which to measure culture or climate is recommended. Several of these instruments exist, and the multitude of available instruments serves as the primary contributor to the problems of reliability and construct validity in the measurement of safety culture. Evans, Glendon, and Creed (2007) and von Thaden and Gibbons (2008) have created instruments whose utility has been demonstrated for flight operations and that could be or have been adapted to maintenance or ground processes as well.

The primary goal of this research was to consolidate research linking safety culture or climate in an aviation organization to safety performance and to evaluate the strength and direction of such a relationship. The paucity of studies that attempted to link these constructs presented a substantial limitation to the present research. Consideration was given to expanding the scope of the literature search to include other industries with similar characteristics to aviation. However, it was decided that to do so failed to address the principal research question of how safety culture or climate affect performance in aviation organizations in particular. Inclusion of research outside of aviation represented another potential source of bias by assuming that any observed correlation between culture and performance in one industry would be generalizable to aviation, an idea contrary to the observations of Ek and Akselsson (2007) and Diaz and Cabrera (1997). These two studies showed significantly different measures of culture when aviation operations were compared to other modes of transportation (shipping) or even other roles within the aviation environment (air traffic control, ground handling).

From a methodological perspective, the meta-analysis was conducted largely as an exploration of the data, however sparse. The term *exploration* is critical here in that it points to the largely experimental inclusion of multivariate results in the meta-analytic review. While Card (2012) provides methods for computing *r* from omnibus tests, the authors were unable to locate or derive a procedure for computing *r* from results of multiple regression. In any event, assuming that the computation of *r* is appropriate as it was done here, it seems logical that each study in the meta-analysis would include the same influence of covariates, an assumption that was infeasible in the current study. If access to basic, bivariate correlation information were available, then it would be possible that the effect sizes of interest could be independently calculated. Again, this identifies a However, this data was not accessible for review, and thus the meta-analysis presented here contains another limitation.

Conclusions and Suggestions for Future Research

This research may itself fall victim to the *file drawer problem*, which can skew meta-analytic results by virtue of the tendency to only publish significant results (Borenstein, et al., 2009; Ellis, 2010). The present research provides only limited insight to the problem simply because the number of studies that related safety culture to safety performance in aviation was so small and the type of available studies presented insurmountable methodological challenges. As such, it was impossible to answer the research question definitively or even to properly evaluate the hypotheses presented here. It is however possible to take away valuable information with respect to future research. Looking ahead, the unification of safety culture measurement is a worthy goal to work toward. The opportunity exists for an institution to champion the widespread use of a single measurement instrument, as was done with LOSA at the University of Texas, to create a more robust data set from which analyses can draw. Worth noting as well is that the relationship between safety culture and safety commitment (shown in Figure 1) should be more closely investigated. Though it would not have increased the number of studies considered in this meta-analysis, a measure of safety commitment should be carefully considered for inclusion in any holistic measure of safety culture or climate. As additional research into the constructs of climate and culture are investigated with respect to aviation safety, clear reporting of effect size will enable future research to expand upon the foundations discussed here to allow a more holistic and inclusive view of the body of knowledge and amalgamate results into a meaningful and systemic representation of the whole. At present, varying constructs, instruments, and analysis techniques present researchers an ever-moving target making empirical investigation of the relationship between positive safety culture and safety performance a yeoman’s task.

References

Baker, R. H. (1991). *Climate survey analysis for aviation maintenance safety* (Master's thesis, Naval Postgraduate School, Monterey, CA). Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA356610

Beus, J. M., Payne, S. C., Bergman, M. E., & Arthur, W. (2010). Safety climate and injuries: An examination of theoretical and empirical relationships. *Journal of Applied Psychology, 95*(4), 713-727. http://dx.doi.org/10.1037/a0019164

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex, UK: John Wiley & Sons. http://dx.doi.org/10.1002/9780470743386

Brittingham, C. (2006). *The relationship between Naval aviation mishaps and squadron maintenance safety climate* (Master's thesis, Naval Postgraduate School, Monterey, USA).

Brown, R. L., & Holmes, H. (1986). The use of a factor-analytic procedure for assessing the validity of an employee safety climate model. *Accident Analysis and Prevention, 18*(6), 455-470. http://dx.doi.org/10.1016/0001-4575(86)90019-9

Burke, M. J., Sarpy, S. A., Tesluk, P. E., & Smith-Crowe, K. (2002). General safety performance: A test of a grounded theoretical model. *Personnel Psychology, 55*(2), 429-457. http://dx.doi.org/10.1111/j.1744-6570.2002.tb00116.x

Card, N. A. (2012). *Applied meta-analysis for social science research.* New York, NY: The Guilford Press.

Clarke, S. (1999). Perceptions of organizational safety: Implications for the development of safety culture. *Journal of Organizational Behaviour, 20*(2), 185-198.

Clarke, S. (2000). Safety culture: Under-specified and overrated? *International Journal of Management Reviews, 2*(1), 65-90. http://dx.doi.org/10.1111/1468-2370.00031

Clarke, S. (2006a). Safety climate in an automobile manufacturing plant. *Personnel Review, 35*(4), 413-430. http://dx.doi.org/10.1108/00483480610670580

Clarke, S. (2006b). The relationship between safety climate and safety performance: A meta-analytic review. *Journal of Occupational Health Psychology, 11*(4), 315-327.

Clarke, S., & Ward, K. (2006). The role of leader influence, tactics, and safety climate in engaging employees' safety participation. *Risk Analysis, 26*(5), 1175-1185. http://dx.doi.org/10.1111/j.1539-6924.2006.00824.x

The Cochrane Collaboration (n.d.). The Cochrane Collaboration. Retrieved from http://www.cochrane.org/

Confederation of British Industry. (1991). *Developing a safety culture.* London, UK: Author.

Cooper, M. D. (2000). Towards a model of safety culture. *Safety Science, 36*(2), 111-136. http://dx.doi.org/10.1016/S0925-7535(00)00035-7

Cox, S., & Flin, R. (1998). Safety culture: Philosopher's stone or man of straw? *Work & Stress, 12*(3), 189-201. http://dx.doi.org/10.1080/02678379808256861

Coyle, I. R., Sleeman, S. D., & Adams, N. (1995). Safety climate. *Journal of Safety Research, 26*(4), 247-254. http://dx.doi.org/10.1016/0022-4375(95)00020-Q

Denison, D. (1996). What is the difference between organizational culture and organizational climate? A native's point of view on a decade of paradigm wars. *Academy of Management Review, 21*(3), 619-654.

Desai, V. M., Roberts, K. H., & Ciavarelli, A. P. (2006). The relationship between safety climate and recent accidents: Behavioural learning and cognitive attributions. *Human Factors, 48*(4), 639-650. http://dx.doi.org/10.1518/001872006779166361

Diaz, R. I., & Cabrera, D. D. (1997). Safety climate and attitude as evaluation measures of organizational safety. *Accident Analysis and Prevention, 29*(5), 643-650. http://dx.doi.org/10.1016/S0001-4575(97)00015-8

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results.* Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511761676

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 34*(1-3), 177-192. http://dx.doi.org/10.1016/S0925-7535(00)00012-6

Fullarton, C., & Stokes, M. (2007). The utility of a workplace injury instrument in prediction of workplace injury. *Accident Analysis & Prevention, 39*(1), 28-37. http://dx.doi.org/10.1016/j.aap.2006.05.015

Glendon, A. I., Clarke, S. G., & McKenna, E. F. (2006). *Human safety and risk management* (2nd ed.). Boca Raton, FL: Taylor & Francis Group.

Gregorich, S. E., Helmreich, R. L., & Wilhelm, J. A. (1990). The structure of cockpit management attitudes. *Journal of Applied Psychology, 75*(6), 682-690. http://dx.doi.org/10.1037/0021-9010.75.6.682

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*(3), 347-358. http://dx.doi.org/10.1037/1076-8998.5.3.347

Guldenmuud, F. W. (2000). The nature of safety culture: A review of theory and research. *Safety Science, 34*(1-3), 215-257. http://dx.doi.org/10.1016/S0925-7535(00)00014-X

Hayes, B. E., Perander, J., Smecko, T., & Trask, J. (1998). Measuring perceptions of workplace safety: Development and validation of the Work Safety Scale. *Journal of Safety Research, 29*(3), 145-161. http://dx.doi.org/10.1016/S0022-4375(98)00011-5

Health and Safety Commission. (1993). *Organising for safety: Third report of the human factors study group of ACSNI.* London, UK: HMSO.

*Helmreich, R. L., Foushee, H. C., Benson, R., & Russini, R. (1986). Cockpit resource management: Exploring the attitude-performance linkage. *Aviation, Space and Environmental Medicine,* (57), 1198-1200.

Helmreich, R. L., & Merittt, A. C. (1998). *Culture at work in aviation and medicine: National, organizational, and professional influences.* Aldershot, UK: Ashgate.

Helmreich, R. L., Merritt, A. C., Sherman, P. J., Gregorich, S. E., & Wiener, E. L. (n.d.). *The Flight Management Attitudes Questionnaire (FMAQ)* (93-4). Austin, TX: The University of Texas.

*Hernandez, A. E. (2001). *Organizational climate and its relationship with aviation maintenance safety* (Master's thesis, Naval Postgraduate School, Monterey, USA).

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557-560. http://dx.doi.org/10.1136/bmj.327.7414.557

Hoffman, 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*(4), 847-869. http://dx.doi.org/10.1111/j.1744-6570.2006.00056.x

Hoffman, D. A., & Stetzer, A. (1996). A cross-level investigation of factors influencing unsafe behaviors and accidents. *Personnel Psychology, 49*(2), 307-339. http://dx.doi.org/10.1111/j.1744-6570.1996.tb01802.x

Human Engineering. (2005). *A review of the safety culture and safety climate literature for the development of the safety culture inspection toolkit* (RR367). Bristol, UK: Author.

IAEA (1991). *Safety Culture* (Safety Series No. 75-INSAG-4). Vienna, Austria: International Atomic Energy Agency.

Kao, L. S., & Thomas, E. J. (2008). Navigating towards improved surgical safety using aviation-based strategies. *Journal of Surgical Research, 145*, 327-335. http://dx.doi.org/10.1016/j.jss.2007.02.020

Kelly, T., & Patankar, M. S. (2004). *Comparison of organizational safety cultures at two aviation organizations*. Paper presented at the Safety Across High-Consequence Industries Conference. St. Louis, MO.

Kennedy, R., & Kirwan, B. (1995). The failure mechanisms of safety culture. In A. Carnino and G. Weimann (Eds.) *Proceedings of the International Topical Meeting on Safety Culture in Nuclear Installations,* (pp.281-290). Vienna, Austria: American Nuclear Society of Austria.

Libuser, C. B. (1994). *Organizational structure and risk mitigation* (Doctoral dissertation). Available from ProQuest: Dissertations and Theses (AAT 9427348).

Makino, M. (2006, March). *Monitoring toward cultivation of safety culture*. Paper presented at the Nuclear Safety Commission International Symposium on Safety Culture. Tokyo, Japan

McDonald, N., Corrigan, S., Daly, C., & Cromie, S. (2000). Safety management systems and safety culture in aircraft maintenance organisations. *Safety Science, 34*(1-3), 151-176. http://dx.doi.org/10.1016/S0925-7535(00)00011-4

Mearns, K., Whitaker, S. M., & Flin, R. (2003). Safety climate, safety management practice and safety performance in offshore environments. *Safety Science, 41*(8), 641-680. http://dx.doi.org/10.1016/S0925-7535(02)00011-5

Neal, A., & Griffin, M. A. (2002). Safety climate and safety behaviour. *Australian Journal of Management, 27*, 67-77. http://dx.doi.org/10.1177/031289620202701S08

Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behavior. *Safety Science, 34*(1-3), 99-109. http://dx.doi.org/10.1016/S0925-7535(00)00008-4

Ostrom, L., Wilhelmsen, C., & Kaplan, B. (1993). Assessing safety culture. *Nuclear Safety, 34*(2), 163-172.

Nieva, V. F., & Sorra, J. (2003). Safety culture assessment: A tool for improving patient safety in healthcare organizations. *Quality & Safety in Healthcare, 12*(Suppl II), ii17-ii23. http://dx.doi.org/10.1136/qhc.12.suppl_2.ii17

*O'Connor, P., Buttrey, S., O'Dea, A., & Kennedy, Q. (2011, January 10). *An assessment of the relationship between safety climate and mishap risk in U.S. Naval aviation* (NPS-OR-11-004). Monterey, CA: Naval Postgraduate School, Department of Operations Research.

O'Connor, P., O'Dea, A., Kennedy, Q., & Buttrey, S. E. (2011). Measuring safety climate in aviation: A review and recommendations for the future. *Safety Science, 49*(2), 128-138. http://dx.doi.org/10.1016/j.ssci.2010.10.001

O'Toole, M. (2002). The relationship between employees' perceptions of safety and organizational culture. *Journal of Safety Research, 33*(2), 231-243. http://dx.doi.org/10.1016/S0022-4375(02)00014-2

Parker, D., Lawrie, M., & Hudson, P. (2006). A framework for understanding the development of organisational safety culture. *Safety Science, 44*(6), 551-562. http://dx.doi.org/10.1016/j.ssci.2005.10.004

Pidgeon, N. (1998). Safety culture: Key theoretical issues. *Work & Stress, 12*(3), 202-216. http://dx.doi.org/10.1080/02678379808256862

Reason, J. (1998). Achieving a safe culture: Theory and practice. *Work & Stress, 12*(3), 293-306. http://dx.doi.org/10.1080/02678379808256868

Reason, J., Parker, D., & Lawton, R. (1998). Organizational controls and safety: The varieties of rule-related behaviour. *Journal of Occupational and Organizational Psychology, 71*(4), 289-304. http://dx.doi.org/10.1111/j.2044-8325.1998.tb00678.x

Rundmo, T. (1996). Associations between risk perception and safety. *Safety Science, 24*(3), 197-209. http://dx.doi.org/10.1016/S0925-7535(97)00038-6

Rundmo, T. (2003). Managers' attitudes towards safety and accident prevention. *Safety Science, 41*(7), 557-574. http://dx.doi.org/10.1016/S0925-7535(01)00091-1

Schein, E. H. (2010). *Organizational Culture and Leadership* (4th ed.). San Francisco, CA: Jossey-Bass.

Seo, D. C., Torabi, M. R., Blair, E. H., & Ellis, N. T. (2004). A cross-validation of safety climate scale using confirmatory factor analytic approach. *Journal of Safety Research, 35*(4), 427-445. http://dx.doi.org/10.1016/j.jsr.2004.04.006

*Sexton, J. B., & Klinect, J. R. (2001). The link between safety attitudes and observed performance in flight operations. In *Proceedings of the Eleventh International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.

Shannon, H. S., Mayr, J., & Haines, T. (1997). Overview of the relationship between organizational and workplace factors and injury rates. *Safety Science, 26*(3), 201-217. http://dx.doi.org/10.1016/S0925-7535(97)00043-X

Silva, S., Lima, M. L., & Baptista, C. (2004). OSCI: An organisational and safety climate inventory. *Safety Science, 41*(3), 205-220. http://dx.doi.org/10.1016/S0925-7535(03)00043-2

Siu, O., Phillips, D. R., & Leung, T. (2004). Safety climate and safety performance among construction workers in Hong Kong: The role of psychological strains as mediators. *Accident Analysis & Prevention, 36*(3), 359-366. http://dx.doi.org/10.1016/S0001-4575(03)00016-2

Smith, A. P., & Wadsworth, E. J. (2009). *Safety culture, advice and performance* (Research Report 09.1). Leicestershire, UK: IOSH.

Smith-Crowe, K., Burke, M. J., & Landis, R. S. (2003). Organizational climate as a moderator of safety knowledge-safety performance relationships. *Journal of Organizational Behavior, 24*(7), 861-876. http://dx.doi.org/10.1002/job.217

Stolzer, A. J., Halford, C. D., & Goglia, J. J. (2008). *Safety management systems in aviation.* Surrey, UK: Ashgate.

Thomas, J. M., Melia, J. L., & Oliver, A. (1999). A cross-validation of a structural equation model of accidents: Organizational and psychological variables as predictors of work safety. *Work & Stress, 13*(1), 49-58. http://dx.doi.org/10.1080/026783799296183

Varonen, U., & Mattila, M. (2000). The safety climate and its relationship to safety practices, safety of the work environment and occupational accidents in eight wood-processing companies. *Accident Analysis & Prevention, 32*(6), 761-769. http://dx.doi.org/10.1016/S0001-4575(99)00129-3

Vinodkumar, M. N., & Bhasi, M. (2009). Safety climate factors and its relationship with accidents and personal attributes in the chemical industry. *Safety Science, 47*(5), 659-667. http://dx.doi.org/10.1016/j.ssci.2008.09.004

von Thaden, T. L., & Gibbons, A. M. (2008). *The Safety Culture Indicator Scale Measurement System (SCISMS)* (Technical Report HFD-08-03/FAA-08-02). Savoy, IL: University of Illinois, Human Factors Division.

Waring, A. (2002). *Safety management systems* (Rev. ed.). London, UK: Chapman & Hall.

Williamson, A. M., Feyer, A., Cairns, D., & Biancotti, D. (1997). The development of a measure of safety climate: The role of safety perceptions and attitudes. *Safety Science, 25*(1-3), 15-27. http://dx.doi.org/10.1016/S0925-7535(97)00020-9

Yule, S. (2003). *Senior management influence on safety performance in the UK and US energy sectors* (Doctoral dissertation, University of Aberdeen, Scotland).

Zhang, H., Wiegmann, D. A., & Von Thaden, T. L. (2002). Safety culture: A concept in chaos? In *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors and Ergonomics Society.

Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology, 65*(1), 96-102.

Zohar, D. (2002). The effects of leadership dimensions, safety climate, and assigned priorities on minor injuries in work groups. *Journal of Organizational Behavior, 23*(1), 75-92. http://dx.doi.org/10.1037/0021-9010.65.1.96

Zohar, D., & Luria, G. (2003). The use of supervisory practices as leverage to improve safety behavior: A cross-level intervention model. *Journal of Safety Research, 34*(5), 567-577. http://dx.doi.org/10.1002/job.130

Zohar, D., & Luria, G. (2004). Climate as a social-cognitive construction of supervisory safety practices: Scripts as proxy of behavior patterns. *Journal of Applied Psychology, 89*(2), 322-333. http://dx.doi.org/10.1016/j.jsr.2003.05.006

*Indicates references included in the meta-analysis.