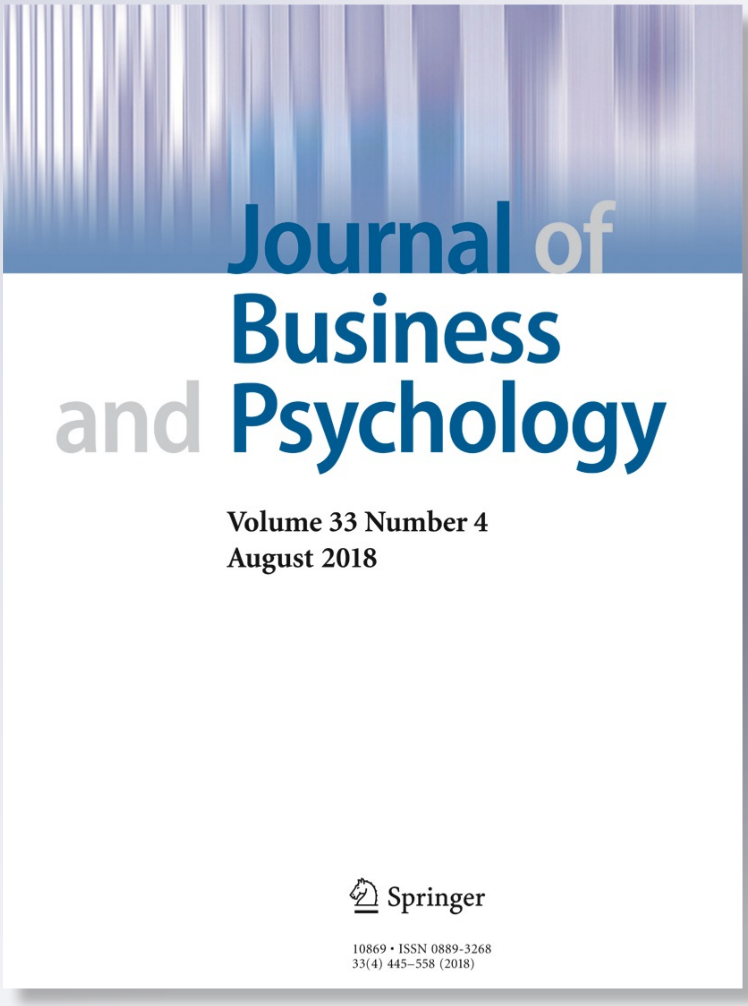
Safety Climate Measurement: an Empirical Test of Context-Specific vs. General Assessments

Nathanael L. Keiser & Stephanie C. Payne

Journal of Business and Psychology

ISSN 0889-3268
Volume 33
Number 4

J Bus Psychol (2018) 33:479-494
DOI 10.1007/s10869-017-9504-y



Journal of Business and Psychology

Volume 33 Number 4
August 2018

Springer

10869 • ISSN 0889-3268
33(4) 445–558 (2018)

Springer

J Bus Psychol (2018) 33:479–494
DOI 10.1007/s10869-017-9504-y

ORIGINAL PAPER

Safety Climate Measurement: an Empirical Test of Context-Specific vs. General Assessments

Nathanael L. Keiser[1] · **Stephanie C. Payne**[1]

Abstract
Purpose Safety climate researchers develop and use both general and industry-specific safety climate measures. Theories about language comprehension suggest that context facilitates meaning; however, the relative value of context-specific safety climate measures in the prediction of safety outcomes is an empirical question that has not been rigorously tested. The purpose of the present study was to provide a rigorous comparison of context-specific vs. general safety climate measures.
Design/Methodology/Approach Seven hundred forty-six university laboratory personnel from five different kinds of research labs (i.e., animal biological, biological, chemical, human subjects/computer, or mechanical/electrical) completed contextualized safety climate measures, a general safety climate measure, and measures of other safety-related constructs.
Findings Measurement equivalence analyses indicated that the general safety climate measure was not equivalent across the five lab types. Hypothesis testing revealed that contextualized information was most helpful when included in safety climate measures for less, rather than more, safety-salient contexts, but overall, there was relatively little difference in the validities for general and context-specific measures.
Implications Results suggest that context has a small influence on how individuals respond to safety climate measures and provide guidance for researchers/practitioners when deciding between using industry-specific or general safety climate measures. It appears most beneficial to use industry-specific measures when examining safety climate in a less-safety-salient context.
Originality/Value This study offers one of the first empirical tests of a contextualized safety climate measure involving a rigorous, unconfounded comparison of five context-specific safety climate measures with a general measure.

Keywords Workplace safety · Safety climate · Industry specific · Contextualization · Laboratory safety

A major tenet of workplace safety research involves effectively predicting and limiting workplace accidents and injuries. Unfortunately, workplace injuries and deaths continue to plague organizations. In 2015, 4836 US workers were killed on the job, and in the same year, over 1.1 million US employees reported nonfatal injuries and illnesses that resulted in lost work days (US Bureau of Labor Statistics, 2015a, b).

There are a number of variables that contribute to workplace safety. This study focuses on safety climate: employees' shared perceptions of the policies, procedures, and practices concerning safety (Zohar, 2003). Given its theoretical importance and empirical promise, multiple measures of safety climate exist in the multidisciplinary safety literature. Some of those measures include industry-specific information, whereas others are general. General measures do not include any industry-specific information making them relevant to employees working in any industry. Industry-specific measures include risks, equipment, and/or procedures that are specific to the industry of interest. Incorporating industry-specific

---

This paper is based on Nathanael Keiser's master's thesis. We thank the committee members Drs. Winfred Arthur, Jr., Mindy Bergman, and M. Sam Mannan for their guidance and support and Xiaohong (Violet) Xu for her assistance with the measurement equivalence analyses. This work was supported by the Mary Kay O'Connor Process Safety Center.

✉ Nathanael L. Keiser
nkeiser@tamu.edu

[1] Department of Psychology, Texas A&M University, College Station, TX 77843-4235, USA

information contextualizes the safety climate measure. Correspondingly, theory and research on contextualization offer an explanation for why industry-specific measures of safety climate might be psychometrically preferable to general measures.

The extent to which industry-specific safety climate measures are more strongly related to safety outcomes compared with a general safety climate measure is an empirical question that has yet to be rigorously tested. Cognitive psychology theories about the role of context in the facilitation of comprehension support the idea that contextual information is beneficial. The present study empirically assesses the difference between context-specific safety climate measures and a general safety climate measure with five samples of university laboratory personnel. Due to the inherent differences between research labs, five laboratory-specific safety climate measures were developed: animal biological, biological, chemical, human subjects/computer, and mechanical/electrical. These five measures will be referred to as context-specific/contextualized measures hereafter.

The *purpose* of the present study is to provide a rigorous empirical comparison of a context-specific safety climate measure with a general safety climate measure. The magnitude of the relationships between safety climate and multiple theoretically relevant and empirically supported outcomes including safety knowledge, compliance, participation, injuries, incidents, and near misses (Christian, Bradley, Wallace, & Burke, 2009) are compared when safety climate is operationalized with a general measure and when it is operationalized with a context-specific measure. The results will begin to inform researchers and practitioners about the utility of developing and administering industry-specific safety climate measures.

Safety Climate and Contextualization

Safety climate is a group-level construct consisting of shared perceptions about safety policies, practices, and procedures (Zohar, 1980). Further, safety climate exists at all levels within an organization and encompasses the formal written policies and procedures as well as the informal unwritten practices that actually take place (Jex, Swanson, & Grubb, 2013). Across multiple theoretical models and frameworks of workplace safety, safety climate is portrayed and empirically supported to be a key predictor of both proximal and distal safety-related outcomes including safety knowledge, safety motivation, safety behavior, and injuries (Christian et al., 2009; Neal & Griffin, 2004).

Measurement of Safety Climate

As a perceptual construct, safety climate is assessed via self-report measures administered to a group of employees and subsequently aggregated to multiple levels (i.e., group, organizational). Safety climate, however, is also conceptualized and assessed at the individual or psychological level (Ostroff, Kinicki, & Muhammad, 2013). Psychological- and organizational-level safety climate measures differ with respect to the theoretical processes through which they emerge and their relationships with behavior and outcomes (Beus, Payne, Bergman, & Arthur, 2010; Kozlowski & Klein, 2000; Ostroff & Bowen, 2000). Indeed, Beus et al. (2010) found that the injury-psychological safety climate relationship ($\rho = -.16$) differed from the injury-organizational safety climate relationship ($\rho = -.29$). This study examines safety climate at the psychological level.

Numerous self-report safety climate measures exist in the safety climate literature. In their recent meta-analysis, Beus et al. (2010) identified 61 unique safety climate measures. In a follow-up systematic review of over 1500 items within these measures (Beus, Payne, & Arthur, 2011), 33 of the 61 measures included at least one industry-specific item (e.g., "Policies regarding not recapping used needles are posted;" Day, 1999, p. 88), whereas 28 measures consisted of only general items (e.g., "A busy situation does not prevent supervisors from intervening if someone acts against safety rules;" Varonen & Mattila, 2000, p. 765). However, the measures differed with respect to the proportion of industry-specific items included with some incorporating all industry-specific items and others including a combination of industry-specific and general items. Industry-specific safety climate measures have been developed for a number of industries, including driving (Wills, Watson, & Biggs, 2006), aviation (Gaba, Singer, Sinaiko, Bowen, & Ciavarelli, 2003), remote electrical/utility workers (Huang et al., 2013), and medical contexts (Gershon et al., 1995).

In a review of 30 years of safety climate research, Zohar (2010) advocated for the development of industry-specific measures of safety climate. He contended that industry-specific measures hone in on those climate perceptions that are especially relevant to a particular industry. As an example, he noted the following item for the transportation industry: "My dispatcher insists that I do not use in-vehicle communication devices while driving (Zohar, 2010, p. 1521)." As this example indicates, industry-specific items frequently incorporate industry-relevant hazards (i.e., in-vehicle communication devices).

Safety Climate Contextualization

Context is fundamental to the study of human behavior. A long-held and well-accepted psychological theory is the premise that behavior is influenced by the interaction between a person and the situation, which can also be described as the context (Lewin, 1951). Context has been defined as "the historical, ethical, political, cultural, environmental, or circumstantial settings or conditions that influence and complicate the consideration of any issues, ideas, artifacts, and events" (Association of American Colleges and Universities, 2010, p. 1). Thus, context encapsulates a variety of situational variables.

Contextualization is most simply defined as *the provision of context*. The contextualization of survey items can be conceptualized on a continuum based on how much context is incorporated into them. At one end of the item-level contextualization continuum are items with little to no context. For example, some personality items do not include contextual information (e.g., "I love excitement"; Johnson, 2014, p. 81). In the middle of the continuum are items which include contextual information but for broad domains (e.g., work, school). In personality research, this is typically done by adding the phrase "at [context]" to the end of the item and/or in the instructions prior to completing a measure (Shaffer & Postlethwaite, 2012). At the other end of the contextualization continuum are items which include contextual information for a specific context or industry (e.g., transportation), job (e.g., welding), or organization (e.g., XYZ Corporation). For example, a climate item for nursing reads, "the nurse manager on this unit makes every effort to see that nurses have the equipment/resources they need to ensure patient/nurse safety" (Hofmann & Mark, 2006, p. 855). The most context-specific items incorporate such unique, context-dependent information that they are not relevant outside of that context (e.g., "When there is a shortage of vehicles, my commander puts the ammunition with the soldiers, though this is against regulations"; Luria, 2008, p. 46). Industry-specific safety climate items tend to fall on the high end of the contextualization continuum.

In much the same way, contextualization can also be conceptualized on a continuum at the measure level based on the proportion of industry-specific items, relative to general items, included in a measure. At one end of the measure-level contextualization continuum are measures that include all general items, and in turn are applicable across contexts (e.g., Beus et al., 2011). At the other end of the continuum are measures that include only context-specific items (e.g., items only applicable to military personnel; Luria, 2008). In the middle of the continuum are measures that include some combination of industry-specific and general items (e.g., Hofmann & Mark, 2006). This study compares safety climate measures on either end of the measure-level continuum; that is, it compares a completely contextualized safety climate measure with a completely general safety climate measure.

Research related to measure contextualization generally indicates that context-specific measures are more predictive of relevant outcomes than general measures. Indeed, in an effort to improve the predictive validity of individual difference and personality measures, researchers have added context or a "frame of reference" to measures of various psychological constructs. The contextualization of personality measures has improved their criterion-related validity as predictors of job performance (Bing, Whanger, Davison, & Van Hook, 2004; Hunthausen, Truxillo, Bauer, & Hammer, 2003; Lievens, De Corte, & Schollaert, 2008; Schmit, Ryan, Stierwalt, & Powell, 1995). In a recent meta-analysis of the frame-of-reference effect, Shaffer and Postlethwaite (2012) found that personality measures altered to cue the work context have greater validity when predicting job performance by an average $d = .29$.

In comparison, there has been limited research on safety climate measure contextualization; a review of the safety climate literature identified only one study that examined the criterion-related validity of an industry-specific safety climate measure. Huang et al. (2013) examined the extent to which truck-driving safety climate measures explained incremental validity in driving behavior and incidents over and above general safety climate measures. Analyses revealed the industry-specific group- and organization-level measures explained an additional 6–11% of variance above and beyond the general group- and organization-level measures. Thus, Huang et al.'s study provides initial evidence that an industry-specific safety climate measure explains additional variance in important safety outcomes beyond a general safety climate measure. The current study is an additional comparison of industry-specific with general safety climate measures that seeks to address three primary research questions: (1) Are context-specific and general safety climate measures psychometrically equivalent within and across contexts? (2) Do context-specific and general safety climate measures differ in their relationships with common safety outcomes? (3) Are criterion-related validity comparisons between context-specific and general safety climate measures similar/different across contexts?

Theory About Context

Cognitive psychology theory and research contend that context facilitates comprehension, and comprehension is an important part of the survey response process. We briefly summarize some of this research to facilitate the development of our hypotheses about why context is likely to influence responses to survey items.

The Survey Response Process

There are a variety of theories concerning the mental steps individuals undergo when responding to survey items. According to Tourangeau and colleagues' (Tourangeau, 1984; Tourangeau & Rasinski, 1988) model, the response process consists of four major steps: (1) comprehension of the item; (2) retrieval of relevant information; (3) use of that information to make required judgments; and (4) selection and reporting of an answer. As noted, the respondent must first comprehend the item which includes linking "key terms to relevant concepts" (Tourangeau, Rips, & Rasinski, 2000, p. 8). In other words, the item must be meaningful to the respondent. It is at this stage that contextual information is likely to facilitate survey responder comprehension.

Contextualized and general items are likely to be interpreted differently simply by virtue of the specificity of the information provided. General items are potentially more ambiguous and consequently open to subjective interpretation. In contrast, contextualized items have less ambiguity facilitating respondent comprehension. For example, a general safety climate item reads, "my co-workers always follow safety procedures" (Beus et al., 2011, p. 16). The number and nature of safety procedures varies across industries. As such, individuals are likely to differ in what they infer the item to mean. Contextualized information addresses the ambiguity of general items by explicitly incorporating more detailed information to facilitate more consistent interpretation across respondents.

Comprehension

Language comprehension refers to the process of deriving meaning from linguistic information by resolving ambiguities (Hunt & Ellis, 2004; MacDonald, Pearlmutter, & Seidenberg, 1994). Cognitive psychologists differentiate between syntactic and lexical processes. Lexical processes involve determining the *contextually appropriate* meaning of individual words (MacDonald et al., 1994). When respondents complete measures, they are more likely to interpret commonly used words and response categories that are not contextually defined. For example, Belson (1981) found that people apply different age cutoffs to "children." Context assists in lexical processing by providing contextual meaning to individual words and sentences, which is evident in language comprehension during daily interactions. In everyday communication, people provide context by giving a more detailed information or history about a circumstance in order for others to better understand a particular situation. Similarly, prompting survey responders with a list of industry-specific safety equipment (e.g., steel-toed boots, harnesses, shields) is likely to facilitate their understanding of what "safety equipment" means and remind them of what this broad term encompasses in their respective context when asked to rate a survey item about safety equipment.

Hypotheses

In summary, general safety climate items are susceptible to lexical ambiguity, potentially leading to differences in interpretation across industries or contexts. Cognitive theories suggest that contextualization of safety climate items may increase survey responders' comprehension during the response process. Safety climate measure contextualization is assessed in this study by comparing how safety climate relates to six safety outcomes when it is operationalized with a general measure compared with a contextualized measure. Safety knowledge, safety behavior, injuries, incidents, and near misses are the primary outcomes examined in this study and, fittingly, are well-established and important safety constructs based on meta-analytic safety models (e.g., Christian et al., 2009).

Hypothesis 1: Relationships between the general safety climate items and the latent safety climate construct will differ across contexts.

Hypothesis 2: A contextualized safety climate measure will have a significantly stronger positive relationship with (a) safety knowledge and (b) behavior than a general safety climate measure.

Hypothesis 3: A contextualized safety climate measure will have a significantly stronger negative relationship with (a) injuries, (b) incidents, and (c) near misses than a general safety climate measure.

Method

Design, Sample, and Procedure

An e-mail message was sent to all faculty, staff, and students at a large southern public university in the USA inviting all levels of laboratory personnel (i.e., undergraduate, graduate, postdoctoral researcher, laboratory manager, research scientist or associate, and principal investigator) to participate in an online survey on laboratory safety. The e-mail message was sent to a distribution list of more than 65,000 faculty, staff, and students; however, a large majority of those on the distribution list were not laboratory personnel. Unfortunately, the university does not keep a record of the total number of laboratory personnel.[1] However, Environmental Health and Safety personnel estimated that they inspected roughly 3900 teaching

---

[1] An open records request was submitted to obtain a list of all university personnel and students who completed laboratory safety training and a list of principal investigators. The lists obtained were outdated and contained numerous duplicates. Further, student records are protected by the Family Educational Rights and Privacy Act. Nevertheless, there were 3738 unique people on these lists and an additional recruitment e-mail was sent directly to these individuals to ensure that all laboratory personnel had an opportunity to participate. E-mails to 387 of the 3738 individuals were returned as undeliverable, and 12 people replied and indicated that they were not laboratory personnel.

and research rooms in 2015. It is unclear how many of those rooms are just for teaching (e.g., a chemistry class). Further, many principal investigators are likely have to more than one room in their laboratory and multiple laboratory personnel.

In an attempt to ensure that only laboratory personnel responded, the recruitment e-mail asked, "Do you currently work or volunteer in a [University Name] research laboratory?" At the end of the survey, respondents were given the option to enter their name and e-mail address in a raffle drawing to win one of five $100 gift cards. The survey was open for a little over a month and one reminder message was sent.

A total of 746 laboratory personnel responded to the survey and provided relatively complete responses. Slightly more respondents were women (329; 53%). On average, respondents were 31 years of age ($SD = 13.24$) and had worked in the laboratory 3.5 years ($SD = 5.95$). Laboratories were staffed with an average of 10 individuals ($SD = 9.56$). A majority of participants were graduate students ($n = 225$; 37%), followed by undergraduate students ($n = 180$; 30%), research scientists ($n = 101$; 17%), postdoctoral researchers ($n = 28$; 5%), laboratory managers ($n = 25$; 4%), principal investigators ($n = 23$; 4%), and research associates ($n = 22$; 4%). Most respondents were personnel from biological laboratories ($n = 219$; 29%), followed by animal biological ($n = 212$; 28%), human subjects/computer ($n = 126$; 17%), chemical ($n = 124$; 17%), and mechanical/electrical ($n = 65$; 9%) laboratories.

Laboratory personnel completed a general safety climate measure and contextualized safety climate measure based on their corresponding laboratory type, allowing for a within-subjects comparison. Laboratory type was determined based on responses to a skip logic question in which personnel indicated if they worked in an animal biological, biological, chemical, mechanical/electrical, human subject, computer, or other laboratory. The general and contextualized measures were counterbalanced.[2] Participants also completed measures of safety-related outcomes. All measures were completed at the individual level because university administrators prohibited collecting identified data that would have permitted aggregation to the laboratory level.

Measures

General Safety Climate An abbreviated version of Beus et al.'s (2011) safety climate measure was used as the general safety climate measure in this study. The nine items used in the current study were identified based on two considerations: (1) conduciveness to contextualization and (2) factor loadings for seven dimensions reported in Beus et al. (2011). The nine-item measure included one item for each dimension and two items from the management commitment to safety and safety equipment and housekeeping dimensions. An additional item was included from the management commitment dimension because of the importance of this dimension to safety climate (Zohar, 2003). Two items from the equipment and housekeeping dimension were included to assess equipment and housekeeping separately.

General items were altered slightly to make them appropriate for the laboratory environment rather than industry (e.g., "principal investigator (PI)/laboratory manager" was substituted for "supervisor"). These slight alterations are not considered to be contextualization, but rather common practice meant only to ensure that individuals were thinking about a research laboratory (rather than industry). All items were responded to on a 5-point agreement scale ($\alpha = .95$).

Contextualized Safety Climate Five contextualized laboratory measures were developed for this study: animal biological, biological, chemical, human subjects/computer, and mechanical/electrical. The nine general safety climate items from Beus et al. (2011) were contextualized by appending laboratory-specific information. Cues applicable to the particular domain and item content follow each item. An example item for each of the five measures appears in the Appendix. In order to create the contextualized items, specific policies, practices and procedures, equipment, issues, and risks were identified for each laboratory type. This information was gleaned from Furr's (2000) *CRC Handbook of Laboratory Safety*, laboratory safety research (Harper & Watt, 2012; National Research Council, 2014), and university manuals and inspection checklists (Michigan State University, 2014; The Ohio State University, 2014; Princeton University, 2014; Texas A&M University, 2009, 2012; Texas Tech University, n.d.; University of Texas at Austin, 2013; West Virginia University, 2012). Additional information was extracted from semi-structured interviews with at least one PI and/or laboratory manager for each type of laboratory, as well as guided tours of each kind of laboratory. Once the measures were developed, they were also reviewed by some of the PIs and their laboratory members for accuracy and completeness. All items were responded to on a 5-point agreement scale (animal biological laboratory $\alpha = .90$; biological laboratory $\alpha = .89$; chemical laboratory $\alpha = .91$; human subjects/computer laboratory $\alpha = .94$; mechanical/electrical laboratory $\alpha = .90$).

Self-report Safety Knowledge, Behavior, and Safety-Related Events Participants completed Griffin and Neal's (2000) measures of safety knowledge, compliance, and participation. Each measure consists of four items responded to on a 5-point agreement scale (safety knowledge $\alpha = .92$; safety compliance $\alpha = .93$; safety participation $\alpha = .89$).

---

[2] A multivariate ANOVA indicated that order of the general and context-specific measures did not have a significant effect on the contextualized safety climate scores nor the general safety climate scores, $F(1, 642) = .10$, $p = .75$, $\eta^2 < .001$.

Respondents also provided individual safety incident data. Three items were used to assess the number of injuries, incidents (unsafe events that did not result in an injury such as equipment damage or a chemical release), and near misses (near incidents that could have resulted in harm to persons or equipment but did not (e.g., improper use of PPE, a lab left unlocked)) each respondent experienced in the last 12 months.

Perceived Job Risk and Biosafety Level Participants reported perceived job risk using a three-item measure from Jermier, Gaines, and McIntosh (1989). These items were responded to on a 5-point scale, ranging from "almost always untrue" to "almost always true" ($\alpha = .86$). Animal biological and biological laboratory personnel were also asked to indicate the biosafety level of their laboratory (level 1–4). According to the US Department of Health and Human Services, "[biosafety] levels are designed in ascending order, by degree of protection provided to personnel, the environment, and the community" (Jermier, et al., 1989, p. 30).

Data Analysis

Hypothesis 1 stated that the relationships between the general safety climate items and latent safety climate construct differ across contexts, which was examined using measurement equivalence analyses (also referred to as measurement invariance) of the various safety measures within and across the five contexts. Measurement equivalence analyses reveal if the same construct is being measured across groups and is established based on a series of sequential tests for configural, metric, scalar, and finally strict invariance (Vandenberg & Lance, 2000). In this study, the configural invariance of one or multiple safety measures is supported if the dimensionality, or factor structure, of the latent construct is the same across contexts. Metric invariance is established if the strength of the relationships between items and the latent construct (i.e., regression slopes) are equivalent. The scalar invariance of one or more measures is supported when the intercepts of the factor loadings are invariant across groups. All three, configural, metric, and scalar invariance, are necessary to provide evidence that the safety construct is measured equivalently within and across laboratory environments (Vandenberg & Lance, 2000). Strict invariance is established if measurement error is the same within or across laboratories; however, measurement error is often unequal and consequently measures rarely meet strict equivalency.

Hypotheses 2 and 3 stated that contextualized safety climate measures display significantly stronger relationships with safety knowledge and behavior, and injuries, incidents, and near misses. These hypotheses were tested using an updated version of Steiger's Z for determining the significance of the difference between dependent correlations (Hoerger, 2013; Steiger, 1980). These hypotheses were further examined using hierarchical multiple regression analyses based on the additional variance explained by contextualized measures above general measures in the prediction of safety knowledge, safety behavior, injuries, incidents, and near misses.

Results

Context Comparisons

An underlying assumption made in this study is the five laboratories examined are significantly different from one another in risks (e.g., biosafety levels 1–4) and safety policies and procedures to warrant different contextualized safety climate measures. Preliminary analyses were conducted to examine the extent to which the five laboratories varied significantly on perceived risk, injuries, incidents, near misses, and for a subset of labs, an objective indicator of risk (biosafety level). ANOVA indicated that the laboratory types varied significantly on perceived risk, $F(4, 645) = 20.35$, $p < .001$, $\eta^2 = .11$, incidents, $F(4, 638) = 4.27$, $p = .002$, $\eta^2 = .03$, and near misses, $F(4, 643) = 3.94$, $p = .004$, $\eta^2 = .02$. Post hoc analyses were conducted using the Games-Howell procedure, which revealed that human subjects/computer laboratory members reported significantly lower perceived risk than respondents from the four other lab types and biological laboratory members reported significantly less-perceived risk than animal biological and chemical laboratory members. Additionally, human subjects/computer laboratory members reported fewer incidents and near misses than animal biological and chemical laboratory respondents (Table 1). Both biological and animal biological labs had median biosafety levels of 2 on a 4-point scale.

Measurement Equivalence

Confirmatory factor analyses using maximum likelihood estimation were conducted in Mplus 7.0 (Muthén & Muthén, 1998–2012) for models of configural, metric, scalar, and strict invariance. Beginning with configural invariance, ten sets of measurement equivalence analyses were conducted. Measurement equivalence of the following five focal variables was assessed across the five labs: (1) general safety climate, (2) contextualized safety climate, (3) safety knowledge, (4) safety participation, and (5) safety compliance. Measurement equivalence between general and contextualized safety climate was also assessed within each of the five labs: (6) animal biological, (7) biological, (8) chemical, (9) human subjects/computer, and (10) mechanical/electrical. For example, the configural invariance test of general safety climate involved loading the nine safety climate items on a single factor and assessing the equivalence of the pattern of free and fixed factor

Table 1 Lab context comparisons

Laboratory type	Perceived risk M (SD)	Injuries M (SD)	Incidents M (SD)	Near misses M (SD)	Biosafety level Mdn
Animal biological	1.86 (0.83) a	.33 (1.12)	.49 (1.40) a	.72 (1.61) a	2.00
Biological	1.60 (0.71) a, b	.18 (0.67)	.26 (0.74)	.58 (1.53)	2.00
Chemical	1.99 (0.94) b	.56 (2.63)	.86 (2.67) b	1.19 (2.80) b	
Human subjects/computer	1.16 (0.45) a, b, c	.15 (0.66)	.10 (0.43) a, b	.27 (0.84) a, b	
Mechanical/electrical	1.82 (0.90) c	.38 (0.97)	.91 (3.48)	.78 (1.80)	

Note. Numbers with the same lowercase letters within a column are significantly different from each other based on Games-Howell post hoc analyses (p <.05, two-tailed)

loadings across labs (Vandenberg & Lance, 2000). The first item's factor loading and intercept was fixed at 1 and 0, respectively; all other aspects of the model across labs were freely estimated (i.e., factor loadings, factor variances, covariances, and means; Vandenberg & Lance, 2000). Subsequent model tests involved constraining the factor loadings (i.e., metric invariance), then the item intercepts (i.e., scalar invariance), and finally the error variances (strict invariance) in the same single-factor model across labs.

Models were compared based on the chi-square difference test (Bollen, 1989). A significant difference indicates that the more restrictive model provides significantly worse model fit. Models were also examined based on the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), and the comparative fit index (CFI). Typical cutoffs for a model with adequate fit are RMSEA less than .06, SRMR less than .08, and CFI greater than .90 (Hu & Bentler, 1999). These fit indices were selected because they are commonly used in confirmatory factor analysis and offer useful information concerning model fit especially when combined (Kline, 2011). However, fit indices are best treated as descriptive indicators, because models with fit indices that pass common thresholds (e.g., Hu & Bentler, 1999) can still provide poor representation of the data (Kline, 2011).

General and Contextualized Safety Climate Across Contexts Fit indices for the configural equivalence model of the general safety climate measure suggested that the configural model provided adequate fit (see Table 2). Fit indices and the chi-square difference ($\Delta\chi^2$ (32) = 28.27, $p > .05$) between the configural and metric model indicated that the metric model was not significantly different from the configural model. These results suggest that relationships between the general safety climate items and latent safety climate construct were equivalent across laboratory types (i.e., metric invariant). The next step involved an examination of the scalar equivalence model and comparison with the metric invariance model. The scalar equivalence model did not provide adequate fit based on two of three fit indices. Moreover, the chi-square difference test was significant ($\Delta\chi^2$ (32) = 68.54, $p < .05$), which indicated that the scalar equivalence model provided significantly worse fit than the metric invariance model. These results suggest that the relationship between the general safety climate items and the latent construct varied across labs, which provides support for hypothesis 1 that context-specific measures are needed. Results for the strict equivalence model are not presented considering the safety climate measure was not scalar invariant.

The equivalence of the five contextualized measures was also assessed considering the same stem was used for each item within the contextualized measures (see Table 2). The configural model provided adequate fit based on two of the three fit indices. Fit indices for the metric invariance model and comparison based on the chi-square difference test, $\Delta\chi^2$ (32) = 34.88, $p > .05$, suggested that the metric equivalence model did not fit appreciably worse than the configural equivalence model. Metric and scalar equivalence model comparisons indicated that the context-specific assessments were not scalar equivalent ($\Delta\chi^2$ (32) = 74.86, $p < .05$). Similar to the general measure, the relationships between contextualized items and the latent safety climate construct differed across lab types.

General and Contextualized Safety Climate Within Contexts Additional measurement equivalence analyses involved a comparison of the general and the contextualized safety climate measures within each type of laboratory (see Table 3). For animal biological and biological labs, the general and corresponding context-specific safety climate measures reached metric equivalence ($\Delta\chi^2$ (8) = 4.03, $p > .05$; $\Delta\chi^2$ (8) = 3.93, $p > .05$). However, fit indices and comparisons of metric and scalar invariance models suggested that the measures for these two labs were not scalar equivalent ($\Delta\chi^2$ (8) = 24.95, $p < .05$; $\Delta\chi^2$ (8) = 33.83, $p < .05$). Information criteria for safety climate measures administered to chemical lab respondents suggested that the configural equivalence model did not provide adequate fit and thus no further models were examined. The safety climate measures for human subjects/computer and mechanical/electrical labs were metric equivalent ($\Delta\chi^2$ (8) = 13.91, $p > .05$; $\Delta\chi^2$ (8) = 8.97, $p > .05$).

Table 2 Measurement equivalence of the study measures across five labs

		χ^2 (*df*)	$\Delta\chi^2$ (Δ*df*)	RMSEA	CFI	SRMR
General safety climate measure	Configural equivalence	297.08* (135)	–	.10	.93	.04
	Metric equivalence	325.35* (167)	28.27 (32)	.09	.94	.08
	Scalar equivalence	393.89* (199)	68.54* (32)	.09	.92	.10
	Strict equivalence	455.11* (235)	–	.09	.91	.14
Contextualized safety climate measures	Configural equivalence[1]	336.45* (135)	–	.11	.90	.06
	Metric equivalence	371.33* (167)	34.88 (32)	.10	.90	.11
	Scalar equivalence	446.19* (199)	74.86* (32)	.10	.88	.12
	Strict equivalence	487.95* (235)	–	.09	.87	.18
Safety knowledge	Configural equivalence	44.94* (10)	–	.16	.96	.03
	Metric equivalence	65.12* (22)	20.18 (12)	.12	.95	.15
	Scalar equivalence	88.25* (34)	23.13 (12)	.11	.94	.16
	Strict equivalence	110.22* (50)	21.97 (16)	.10	.93	.14
Safety participation	Configural equivalence	45.01* (10)	–	.17	.96	.03
	Metric equivalence	62.59* (22)	17.58 (12)	.12	.95	.11
	Scalar equivalence	78.58* (34)	15.99 (12)	.10	.95	.12
	Strict equivalence	86.90* (50)	8.32 (16)	.08	.96	.16
Safety compliance	Configural equivalence	8.88 (10)	–	.00	1.00	.01
	Metric equivalence	21.95 (22)	13.07 (12)	.00	1.00	.14
	Scalar equivalence	31.24 (34)	9.29 (12)	.00	1.00	.15
	Strict equivalence	53.08 (50)	21.84 (16)	.02	1.00	.18

*p <.05, two-tailed

Table 3 Measurement equivalence of the general safety climate measure and contextualized safety climate measures within five labs

		χ^2 (*df*)	$\Delta\chi^2$ (Δ*df*)	RMSEA	CFI	SRMR
Animal biological ($n = 212$)	Configural equivalence	141.41* (54)	–	.09	.92	.05
	Metric equivalence	145.44* (62)	4.03 (8)	.09	.93	.07
	Scalar equivalence	170.39* (70)	24.95* (8)	.09	.91	.08
	Strict equivalence	194.80* (79)	–	.09	.90	.14
Biological ($n = 219$)	Configural equivalence	126.62* (54)	–	.08	.93	.05
	Metric equivalence	130.55* (62)	3.93 (8)	.08	.94	.07
	Scalar equivalence	164.38* (70)	33.83* (8)	.08	.91	.09
	Strict equivalence	189.39* (79)	–	.09	.90	.16
Chemical ($n = 124$)	Configural equivalence	123.93* (54)	–	.11	.89	.06
	Metric equivalence	131.66* (62)	–	.10	.89	.08
	Scalar equivalence	155.43* (70)	–	.11	.87	.10
	Strict equivalence	172.69* (79)	–	.10	.85	.14
Human subjects/computer ($n = 126$)	Configural equivalence	129.58* (54)	–	.14	.91	.04
	Metric equivalence	143.49* (62)	13.91 (8)	.13	.91	.07
	Scalar equivalence	159.78* (70)	16.29 (8)	.12	.90	.07
	Strict equivalence	190.62* (79)	30.84* (9)	.11	.89	.10
Mechanical/electrical ($n = 65$)	Configural equivalence	97.64* (54)	–	.12	.91	.06
	Metric equivalence	106.61* (62)	8.97 (8)	.11	.91	.10
	Scalar equivalence	114.08* (70)	7.47 (8)	.11	.91	.11
	Strict equivalence	121.91* (79)	7.83 (9)	.10	.91	.13

*p < .05, two-tailed

Fit indices and the chi-square difference test ($\Delta\chi^2$ (8) = 16.29, $p > .05$; $\Delta\chi^2$ (8) = 7.47, $p > .05$) also indicated that the safety climate measures for these labs were scalar equivalent. Comparisons of scalar and strict equivalence models supported the strict equivalence for mechanical/electrical labs ($\Delta\chi^2$ (9) = 7.83, $p > .05$), whereas the measures for human subjects/computer labs did not reach strict equivalency ($\Delta\chi^2$ (9) = 30.84, $p < .05$). In sum, the general measure and corresponding context-specific safety climate measures for animal biological, biological, and chemical labs were not equivalent, but the general and corresponding context-specific measures were equivalent for the human subjects/computer and mechanical/electrical labs. These results provide partial support for hypothesis 1.

Safety Knowledge, Participation, and Compliance Across Contexts The measurement equivalence of safety knowledge, participation, and compliance was also assessed across the five labs (see Table 2). Fit indices and comparisons of configural and metric equivalent models suggested that all three measures were metric equivalent ($\Delta\chi^2$ (12) = 20.18, $p > .05$; $\Delta\chi^2$ (12) = 20.58, $p > .05$; $\Delta\chi^2$ (12) = 13.07, $p > .05$). Safety knowledge and behavior were also scalar equivalent based on information criteria and chi-square difference tests comparing metric and scalar invariance models ($\Delta\chi^2$ (12) = 23.13, $p > .05$; $\Delta\chi^2$ (12) = 15.99, $p > .05$; $\Delta\chi^2$ (12) = 9.29, $p > .05$). Comparisons of scalar and strict invariance models for these measures indicated that they were also strict equivalent ($\Delta\chi^2$ (16) = 21.97, $p > .05$; $\Delta\chi^2$ (16) = 8.32, $p > .05$; $\Delta\chi^2$ (16) = 21.84, $p > .05$). These results together suggest that these measures assessed safety knowledge, participation, and compliance equivalently across the different types of labs.

Correlation Comparisons

Descriptive statistics and intercorrelations of study variables are reported in Table 4. Table 5 presents the correlations between safety climate and its outcomes, facilitating a comparison of results obtained with contextualized and general safety climate measures. Hypotheses 2 and 3 predicted the contextualized measures would have stronger relationships with various safety-related outcomes than the general safety climate measure. These hypotheses were not supported for the animal biological, biological, chemical, and mechanical/electrical measures (see Table 5). None of the comparisons between these contextualized measures and the general measure in their relationships with the six outcomes were significantly different. Contrary to hypothesis 2, the contextualized biological measure had a significantly weaker relationship with safety compliance compared with the general measure ($r = .52$ vs. $r = .61$), $Z(186) = -2.68$, $p = .007$.

Results for the human subjects/computer laboratory measure were more supportive of the hypothesized differences. Hypothesis 2 was partially supported for the human subjects/computer measure. Compared with the general safety climate measure, the contextualized safety climate measure had a significantly stronger relationship with safety knowledge ($r = .62$ vs. $r = .43$), $Z(106) = 2.89$, $p = .004$. Correlation comparisons, however, with safety participation ($r = .59$ vs. $r = .47$), $Z(106) = 1.82$, $p = .07$, and compliance ($r = .64$ vs. $r = .53$), $Z(106) = 1.76$, $p = .08$ were not statistically different.

Hypothesis 3 was also partially supported for the human subjects/computer laboratory measure. Whereas the difference in the correlations between contextualized and general safety climate with incidents ($r = -.42$ vs. $r = -.29$), $Z(106) = -1.74$, $p = .08$ and with near misses was not statistically significant ($r = -.36$, vs. $r = -.24$), $Z(106) = -1.57$, $p = .12$, the difference between the correlations was significant for injuries ($r = -.34$ vs. $r = -.13$), $Z(106) = -2.70$, $p = .007$. Thus, contextualization only appears to be making a criterion-related validity difference for the human subjects/computer measure.

As indicated by its title, respondents from two types of labs responded to this measure and fortunately, personnel indicated which of the two labs (human subjects or computer) they worked in. This presented a unique opportunity to determine if contextualization was strengthening validities for both groups equally (see Table 5). Again, Steiger's Z was used to compare the contextualized safety climate correlations with the general safety climate correlations this time differentiating between human subjects and computer lab respondents. For human subject respondents, most of the safety climate correlations were stronger for the contextualized measure but not significantly different. In comparison, the contextualized computer lab safety climate measure had a significantly stronger relationship with safety knowledge ($r = .63$ vs. $r = .41$), $Z(52) = 2.34$, $p = .02$, safety compliance ($r = .64$ vs. $r = .43$), $Z(53) = 2.28$, $p = .02$, and injuries ($r = -.53$ vs. $r = -.30$), $Z(53) = -2.31$, $p = .02$, compared with the general measure. Thus, the differences observed when testing hypotheses 2 and 3 appear to be primarily driven by the stronger criterion-related validities of the contextualized measure for the computer lab.

Additional Analyses

Contextualized and general safety climate measures were also compared via hierarchical multiple regression (Table 6). The results of these analyses reflected the correlational comparisons, such that the human subjects/computer contextualized safety climate measure accounted for the greatest incremental validity above the general measure in the prediction of safety knowledge (20%; $F(1, 102) = 33.18$, $p < .001$), participation (14%; $F(1, 103) = 22.65$, $p < .001$), and compliance (16%; $F(1, 102) = 29.18$, $p < .001$), and injuries (11%; F(1,

Table 4 Descriptive statistics and intercorrelations

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Safety knowledge	4.46	0.58	(.92)											
2. Safety participation	4.12	0.69	.63*	(.89)										
3. Safety compliance	4.40	0.63	.75*	.62*	(.93)									
4. Injuries	.30	1.34	−.08	−.05	–	–								
5. Incidents	.45	1.74	−.04	−.03	−.07	.49*	–							
6. Near misses	.68	1.78	−.04	−.06	−.18*	.39*	.42*	–						
7. General safety climate	4.16	0.76	.47*	.52*	.54*	−.16*	−.12*	−.21*	(.95)					
8. Animal biological[a]	4.25	0.64	.44*	.45*	.47*	−.15*	−.14	−.22*	.80*	(.90)				
9. Biological[a]	4.27	0.57	.41*	.48*	.52*	−.28*	−.21*	−.35*	.84*	–	(.89)			
10. Chemical[a]	4.10	0.69	.52*	.49*	.55*	−.21*	−.21*	−.31*	.86*	–	–	(.91)		
11. Human subjects/computer[a]	4.02	0.81	.62*	.59*	.64*	−.34*	−.42*	−.36*	.66*	–	–	–	(.94)	
12. Mechanical/electrical[a]	4.11	0.57	.64*	.63*	.66*	−.01	−.08	−.12	.70*	–	–	–	–	(.90)

Notes. Total $n = 746$; reliabilities (coefficient alphas) appear on the diagonal

*$p \le .05$

[a] Contextualized safety climate measure (animal biological $n = 212$; biological $n = 219$; chemical $n = 124$; human subjects/computer $n = 126$; mechanical/electrical $n = 65$)

103) = 13.14, $p < .001$), incidents (9%; $F(1, 101) = 11.08$, $p = .001$), and near misses (7%; $F(1, 103) = 7.92$, $p = .006$). In comparison, contextualized measures for the other lab types combined (animal biological, biological, chemical, and mechanical/electrical) explained a substantially smaller proportion of additional variance in the prediction of safety knowledge (2%; $F(1, 523) = 12.05$, $p = .001$), participation (1%; $F(1, 521) = 6.52$, $p = .01$), and compliance (1%; $F(1, 525) = 11.17$, $p = .001$), and injuries (0.4%; $F(1, 522) = 1.94$, $p = .16$), incidents (1%; $F(1, 516) = 3.65$, $p = .06$), and near misses (2%; $F(1, 519) = 11.29$, $p = .001$).

Discussion

Industry-specific items are often included in safety climate measures (Beus et al., 2010; Zohar, 2003, 2010); however, the extent to which this measurement approach results in criterion-related validity gains warranted a rigorous empirical examination. The purpose of this study was to compare the criterion-related validity of a contextualized safety climate measure with a general safety climate measure for six safety-related outcomes. Contextualized safety climate measures were developed for and administered to lab personnel who

Table 5 Correlations between safety climate and six safety outcomes by lab

Laboratory type	Contextualized vs. general	Knowledge	Participation	Compliance	Injuries	Incidents	Near misses
Animal biological ($n = 212$)	Contextualized	.44	.45	.47	−.15	−.14	−.22
	General	.41	.51	.45	−.20	−.10	−.15
Biological ($n = 219$)	Contextualized	.41	.48	.52*	−.28	−.21	−.35
	General	.47	.53	.61*	−.24	−.16	−.28
Chemical ($n = 124$)	Contextualized	.52	.49	.55	−.21	−.21	−.31
	General	.52	.51	.57	−.17	−.17	−.33
Human subjects/computer ($n = 126$)	Contextualized	.62*	.59	.64	−.34*	−.42	−.36
	General	.43*	.47	.53	−.13*	−.29	−.24
Mechanical/electrical ($n = 65$)	Contextualized	.64	.63	.66	−.01	−.08	−.12
	General	.67	.71	.73	−.12	−.12	−.10
Human subjects ($n = 65$)	Contextualized	.61	.54	.64	.05	−.25	−.24
	General	.45	.44	.64	.07	−.17	−.18
Computer ($n = 61$)	Contextualized	.63*	.63	.64*	−.53*	−.48	−.50
	General	.41*	.51	.43*	−.30*	−.37	−.35

*$p < .05$, two-tailed

Table 6 Hierarchical multiple regression analyses

Laboratory type	Safety climate measure	Criteria											
		Knowledge		Participation		Compliance		Injuries		Incidents		Near misses	
		R^2	ΔR^2	R^2	ΔR^2	R^2	ΔR^2	R^2	ΔR^2	R^2	ΔR^2	R^2	ΔR^2
Animal biological (*n* = 212)	General	.17*	–	.26*	–	.20*	–	.04*	–	.01	–	.02*	–
	G + C	.20*	.03*	.26*	.01	.23*	.03*	.04*	.00	.02	.01	.05*	.02*
Biological (*n* = 219)	General	.22*	–	.28*	–	.36*	–	.06*	–	.03*	–	.09*	–
	G + C	.22*	.00	.28*	.00	.36*	.00	.08*	.02	.04*	.02	.12*	.04*
Chemical (*n* = 124)	General	.27*	–	.25*	–	.34*	–	.03	–	.03	–	.11*	–
	G + C	.29*	.02	.26*	.01	.35*	.01	.05	.02	.05	.02	.11*	.00
Human subjects/computer (*n* = 126)	General	.19*	–	.22*	–	.28*	–	.02	–	.08*	–	.06*	–
	G + C	.39*	.20*	.35*	.14*	.45*	.16*	.13*	.11*	.18*	.09*	.13*	.07*
Mechanical/electrical (*n* = 65)	General	.44*	–	.51*	–	.53*	–	.01	–	.01	–	.01	–
	G + C	.50*	.06*	.54*	.04*	.57*	.04*	.03	.01	.01	.00	.02	.01

G + C general safety climate measure combined with contextualized safety climate measure

*$p < .05$, two-tailed

worked in five different kinds of laboratories (animal biological, biological, chemical, human subjects/computer, and mechanical/electrical). Thus, the extent to which contextualization made a difference was assessed for five different samples of laboratory personnel.

Initial analyses supported the assumption that the five laboratories present unique contexts. The labs differed significantly on perceptions of risk and self-reported incidents and near misses. In addition, measurement equivalence analyses indicated that the general and context-specific safety climate measures were nonequivalent across contexts. Further equivalence analyses of the general safety climate measure and context-specific measures within each lab revealed that these two measures were equivalent for human subjects/computer and mechanical/electrical labs but not equivalent for animal biological, biological, and chemical personnel. In comparison, the safety knowledge, participation, and compliance measures were equivalent across labs. These results provide mixed support for the assertions of this study. Specifically, they support the use of contextualized measures considering respondents across labs interpreted the general measure differently and the unique information included in each contextualized measure contributed to measurement nonequivalence. The within-lab equivalence analyses, however, were less supportive of the contextualized approach taken in this study considering animal biological, biological, and chemical personnel interpreted the associated context-specific safety climate measure differently.

When the contextualized and general safety climate correlations with six safety-related outcomes were compared, a majority were statistically equivalent, especially for the animal biological, chemical, and mechanical/electrical labs. That is, the provision of context-specific information did not lead to significant improvement or decrement in criterion-related validities for four out of the five contextualized safety climate measures. Contextualization appeared to be useful for human subjects and computer laboratories. Indeed, the contextualized human subjects/computer measure related more strongly to all six predictors than the general measure, and this difference was statistically significant for injuries. Upon probing these relationships further, it was revealed that contextualization was most beneficial for computer compared with human subject laboratories based on a comparison of contextualized vs. general safety climate correlations with safety knowledge, compliance, and injuries. Thus, contextualization appears to enhance the criterion-related validity of safety climate measures for less-safety-salient contexts.

Contrary to Huang et al. (2013) who found evidence to support the incremental validity of industry-specific measures over general measures, four of the five contextualized measures explained minimal to no additional variance over and above the general measure in the prediction of the six safety outcomes. In contrast, the human subjects/computer measure explained substantially larger additional variance above the general measure in the prediction of the six other safety constructs, compared with the four other context-specific measures.

Theoretical Implications

This study contributes to the industry-specific safety climate literature by providing a preliminary theoretical explanation for the potential utility of contextualized measurement. Consistent with cognitive theories of comprehension (MacDonald et al., 1994), contextual information was valuable for human subjects/

computer personnel based on a comparison of criterion-related validities between the contextualized and general safety climate measure. Contextualized items might have led to stronger relationships because they were more interpretable to human subjects/computer respondents. Further research is necessary to directly test if contextualized information facilitates comprehension. Simply asking respondents if the contextualized items were more comprehensible would provide some initial information about the viability of this theoretical explanation.

The usefulness of contextualized information in safety climate measures might depend on the inherent risk and salience of safety in that context, with contextualization helping more so in less-safety-salient contexts. In this study, five types of laboratories were identified and differentiated in part because they varied on inherent risk. Animal biological, biological, chemical, and mechanical/electrical laboratories utilize equipment and materials that can cause bodily harm or even death (DeRoos, 1977; Furr, 2000). The contextualized safety climate measures developed for these laboratories incorporated some of these risks, including chemical and biohazards, infectious waste, sources of radiation, and dangerous equipment (e.g., power and machine tools, incinerators, lasers, soldering irons, etc.). Most laboratory personnel in these labs are required to complete extensive safety training dealing with general and laboratory-specific risks, although training varies across institutions (National Research Council, 2014). Perhaps the risks are quite obvious to these personnel and contextualization of measures operationalized in this study as an elaboration of context-specific risks and procedures is not necessary for respondent comprehension.

The risks associated with human subjects and computer laboratories are minor in comparison with the other labs. Some of the risks identified for the contextualized human subjects/computer measure include electrical, tripping, and fall hazards. Overall, there is less-safety training for human subjects and computer laboratory personnel. Regulations and training mainly focus on reducing risks for human subject participants, rather than laboratory members (Protection of Human Subjects, 2009).

Thus, contextualization appears to be particularly useful for human subjects and computer personnel, perhaps because they tend to be less experienced with risk and safety regulations and training. A general safety climate measure may be deficient for human subjects and computer laboratory personnel because respondents are unsure of what constitutes a risk in their laboratory and are more likely to question the item's meaning. In turn, they respond based on relatively few (if any) experiences. Presumably, risks in computer labs are even less prevalent than risks in human subject labs. Fittingly, the contextualized measure had the greatest utility for computer labs, arguably the least safety-salient context examined in this study. In comparison, personnel from the other laboratories might have little difficulty interpreting the meaning of general items because safety is an integral aspect of their work duties and training. After all, there were significant differences in perceived risk, self-reported incidents, and self-reported near misses between the laboratories. This explanation relies on some assumptions about differences across labs in training and awareness of risks that were not directly examined in this study but might be a worthwhile endeavor in future research.

This study also contributes more broadly to contextualization of measures of other psychological constructs (see also Shaffer & Postlethwaite, 2012). The results, however, raise questions about the application of the frame-of-reference effect beyond personality measures and the effectiveness of contextualization in other domains. Moreover, results suggest that the effect of contextualization might not be linear. That is, contextualization in this study consisted of more detailed information than simply adding "at work" to safety climate measures. However, the addition of more detailed information did not subsequently lead to strong effects for the context-specific measures, which is at odds with personality research and the contextualization continuum. Definitive conclusions about the application of the frame-of-reference effect for measures of other constructs and tenets inherent in the contextualization continuum warrant further research.

Practical Implications

A primary practical implication of this study is that general and contextualized safety climate measures appear to be equally valid. These results in combination with the practical advantages of each measure provide some guidance for researchers and practitioners when deciding between using a general or contextualized approach. The practical advantages of general measures result from their broad applicability. That is, general measures by their very nature are applicable to people in different contexts. This allows for easy comparison, use, and analysis of measures. For safety climate in particular, general measures are easier to use because the items do not need to be modified or new measures do not need to be developed. General measures permit normative and quantitative comparisons across organizations (i.e., benchmarking) because of the commonality of items within and across industries. General measures apply to a variety of industries and thus do not lead to a proliferation of specific assessments, some of which incorporate construct-irrelevant information (e.g., risk, fatalism, job security; Beus et al., 2010).

Contextualized or industry-specific safety climate measures also have some notable advantages. The results of this study indicated that it appears most beneficial to include context or industry-specific information in safety climate measures when examining safety climate in a less-safety-salient context. Administering industry-specific measures in these contexts might also increase respondent awareness of hazards and risks in their environment. Moreover, one notable reason for the use of contextualized measures in the applied sector is that they might generate more targeted and useful organizational feedback.

That being said, contextualized measures of safety climate can be cumbersome to develop because it is difficult to identify all possible risks, equipment, and procedures.

Limitations and Future Directions

The contributions of this study must be couched within its limitations. First, university administrators' concerns about respondent and PI identity imposed significant restrictions on the design of the study preventing the collection of participant identification information that would have facilitated aggregating safety climate responses within laboratories. Consequently, psychological safety climate rather than group-level safety climate was examined. It remains to be seen if similar results would emerge at the group level of analysis. Study design restrictions also prevented linking survey responses to organizational records of injuries and incidents, limiting the assessment of these constructs to self-report measures. On the positive side, self-report anonymous data are less likely to suffer from the underreporting that plague organizational records.

Likewise, the university was unable to provide a list of the total number of laboratory personnel working at the time of the study. In turn, it is difficult to estimate how many laboratory personnel chose not to respond to the survey. Without more information, the representativeness of the sample is unclear and differences between respondents and non-respondents cannot be tested. The number of principal investigators in the sample is especially small. Nevertheless, relationships among safety constructs in this study are consistent with theoretical propositions and empirical evidence in the workplace safety literature (Christian et al., 2009).

Additionally, safety climate measure correlations with outcomes are likely to be inflated by common method variance (Podsakoff, MacKenzie, & Podsakoff, 2012); however, the magnitude of this inflation is likely to be equivalent across the general and contextualized measures. Correspondingly, common method variance is not a serious concern, because it does not alter our ability to confidently compare correlations. Nevertheless, future efforts to examine contextualization should try to avoid measuring all variables using self-reports in order to determine the magnitude of the contextualization effect independent of common method bias.

Many of the correlations for the contextualized measures were stronger than the generalized measure, but these differences often did not meet the statistical threshold of significance. The sample sizes for each laboratory tended to be small, resulting in less than sufficient power to detect differences. However, it is important to note that even small improvements in workplace safety can have meaningful practical implications for employee health and welfare.

Future research may be directed at alternative ways of contextualizing measures in an effort to find the best way to do so. Indeed, the measurement equivalence results indicated that respondents from animal biological, biological, and chemical labs interpreted the context-specific measures differently, suggesting that future research might involve a different approach at contextualization as a means of improving the interpretation of these measures. The approach taken in the current study was maximizing internal validity by limiting the difference between the contextualized and the general measure to the provision of context, facilitating a rigorous comparison of the measures. This approach, however, differs from typical industry-specific safety climate measures, which tend to be developed independent of general measures and include a mix of industry-specific and general items. The approach taken in this study also involved adding examples to each item, which added considerable text. This may have influenced how closely respondents read the items and their thoroughness in providing a response. Going forward, researchers may want to compare safety climate measures with different degrees of specificity (e.g., organization-specific and site-specific) and/or industry-specific measures with a combination of context-specific and general items.

Finally, the focus of this study was safety climate and the effectiveness of contextualization based on a comparison of criterion-related validities for context-specific vs. general measures. Accordingly, the safety measures used for comparison purposes (i.e., safety knowledge and behavior) were not context specific. However, the theoretical arguments for contextualized safety climate measures might also apply to the measures of other safety constructs. Measures of other safety constructs in this study were invariant across contexts, which suggests that contextualized measures for these contexts might not be necessary. Nevertheless, an interesting avenue for future research is applying and examining contextualization in measures of various safety constructs.

Conclusion

Safety climate experts advocate using industry-specific measures when assessing safety climate (Zohar, 2010). A recent examination revealed that two industry-specific (trucking) safety climate measures accounted for significant incremental validity in relevant outcomes above general safety climate measures (Huang et al., 2013). The present study extends this research by providing a rigorous comparison of context-specific safety climate measures to a general safety climate measure using five samples of research laboratory personnel. Results showed that the context-specific measure had equivalent and sometimes stronger relationships with six safety-related outcomes. Context made the biggest difference in less-safety-salient contexts.

Appendix

Table 7 Sample safety climate items

Lab type	Item
General safety climate: administered across labs	Equipment in my lab is checked to make sure it is free of faults.
Context-specific safety climate: animal biological laboratory	Equipment in my lab is checked to make sure it is free of faults (a list of some of the equipment and furniture that is likely to be in your lab is below. Shortened list below). • Biosafety cabinets • Centrifuges • Autoclaves • Animal cages • Cage washers • Incinerators • Fume hoods • Power and machine tools • X-ray equipment • Burners and hot plates • Refrigerators/freezers • Compressed gas cylinders
Context-specific safety climate: biological laboratory	Equipment in my lab is checked to make sure it is free of faults (a list of some of the equipment and furniture that is likely to be in your lab is below. Shortened list below). • Biosafety cabinets • Centrifuges • Autoclaves • Fume hoods • Lasers • Hydraulically or pneumatically driven equipment • Power and machine tools • X-ray equipment • Exercise equipment • Burners and hot plates • Refrigerators/freezers • Compressed gas cylinders
Context-specific safety climate: human subjects/computer laboratory	Equipment in my lab is checked to make sure it is free of faults (a list of some of the equipment and furniture that is likely to be in your lab is below). • Wiring • Extension cords • Computers • Stepladders • Copiers • Microwaves • Furniture including file cabinets, shelves, desks, and chairs • Office supplies
Context-specific safety climate: mechanical/electrical laboratory	Equipment in my lab is checked to make sure it is free of faults (a list of some of the equipment and furniture that is likely to be in your lab is below. Shortened list below). • Hand tools • Power tools • Hydraulically or pneumatically driven equipment • Batteries, cells, capacitors, etc. • Soldering irons • Electrical circuits • Electrical conductors • Lasers • Chemicals used for cleaning • Ladders • Wiring and cords

References

Association of American Colleges and Universities (2010). *Critical thinking VALUE rubric*. Retrieved from https://aacu.org/.

Belson, W. A. (1981). *The design and understanding of survey questions*. Aldershot, England: Gower.

Beus, J. M., Payne, S. C., & Arthur, W. Jr (2011). *The initial validation of a universal measure of safety climate.* Paper presented at the 26th annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

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

Bing, M. N., Whanger, J. C., Davison, H. K., & Van Hook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology, 89*, 150–157. doi:10.1037/0021-9010.89.1.150.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

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. doi:10.1037/a0016172.

Day, B. T. (1999). *Using stages of change to examine fear, threat, efficacy, and safety climate perceptions in health care workers who routinely handle needles and sharps*. Unpublished doctoral dissertation, West Virginia University.

DeRoos, R. L. (1977). Environmental health and safety in the academic setting. *American Journal of Public Health, 67*, 851–854. doi:10.2105/AJPH.67.9.851.

Furr, A. K. (2000). *CRC handbook of laboratory safety*. Boca Raton: CRC press.

Gaba, D. M., Singer, S. J., Sinaiko, A. D., Bowen, J. D., & Ciavarelli, A. P. (2003). Differences in safety climate between hospital personnel and naval aviators. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*, 173–185. doi:10.1518/hfes.45.2.175.27238.

Gershon, R. R., Vlahov, D., Felknor, S. A., Vesley, D., Johnson, P. C., Delcios, G. L., & Murphy, L. R. (1995). Compliance with universal precautions among health care workers at three regional hospitals. *American Journal of Infection Control, 23*, 225–236. doi:10.1016/0196-6553(95)90067-5.

Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance. *Journal of Occupational Health Psychology, 5*, 347–358. doi:10.1037/1076-8998.5.3.347.

Harper, L., & Watt, F. (2012). *Laboratory safety culture survey*. Retrieved from http://www.bioraft.com.

Hoerger, M. (2013). Z_H: An updated version of Steiger's Z and web-based calculator for testing the statistical significance of the difference between dependent correlations. Retrieved from http://www.psychmike.com/dependent_correlations.php.

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.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. doi:10.1080/10705519909540118.

Huang, Y. H., Zohar, D., Robertson, M. M., Garabet, A., Lee, J., & Murphy, L. A. (2013). Development and validation of safety climate scales for lone workers using truck drivers as exemplar. *Transportation Research Part F: Traffic Psychology and Behaviour, 17*, 5–19. doi:10.1016/j.trf.2012.08.011.

Hunt, R. R., & Ellis, H. C. (2004). *Fundamentals of cognitive psychology*. New York: McGraw-Hill.

Hunthausen, J. M., Truxillo, D. M., Bauer, T. N., & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology, 88*, 545–551. doi:10.1037/0021-9010.88.3.545.

Jermier, J. M., Gaines, J., & McIntosh, N. J. (1989). Reactions to physically dangerous work: A conceptual and empirical analysis. *Journal of Organizational Behavior, 10*, 15–33.

Jex, S. M., Swanson, N., & Grubb, P. (2013). Healthy workplaces. In N. Schmitt, S. Highhouse, & I. Weiner (Eds.), *Handbook of psychology: Industrial and organizational psychology* (pp. 615–642). Hoboken: John Wiley & Sons Inc..

Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. doi:10.1016/j.jrp.2014.05.003.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York City: The Guilford Press.

Kozlowski, S. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco: Jossey-Bass.

Lewin, K. (1951). Formalization and progress in psychology. In D. Cartwright (Ed.), *Field theory in social science*. New York: Harper.

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology, 93*, 268–279. doi:10.1037/0021-9010.93.2.268.

Luria, G. (2008). Climate strength—how leaders form consensus. *The Leadership Quarterly, 19*, 42–53. doi:10.1016/j.leaqua.2007.12.004.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703. doi:10.1037/0033-295X.101.4.676.

Michigan State University (2014). *Biological safety manual*. Retrieved from http://www.orcbs.msu.edu/.

Muthén, L.K., & Muthén, B.O (1998–2012). *Mplus User's Guide* (7th edition). Los Angeles, CA: Muthén & Muthén.

National Research Council. (2014). *Safe science: Promoting a culture of safety in academic chemical research*. Washington, DC: The National Academies Press.

Neal, A., & Griffin, M. A. (2004). Safety climate and safety at work. In J. Barling, M. R. Frone, J. Barling, & M. R. Frone (Eds.), *The psychology of workplace safety* (pp. 15–34). Washington, DC: American Psychological Association.

Ostroff, C., & Bowen, D. E. (2000). Moving HR to a higher level: HR practices and organizational effectiveness. In K. J. Klein & S. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 211–266). San Francisco: Jossey-Bass.

Ostroff, C., Kinicki, A. J., & Muhammad, R. S. (2013). Organizational culture and climate. In N. W. Schmitt, S. Highhouse, & I. B. Weiner (Eds.), *Handbook of psychology: Industrial and organizational psychology* (pp. 643–676). Hoboken: John Wiley & Sons Inc..

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569. doi:10.1146/annurev-psych-120710-100452.

Princeton University (2014). Laboratory safety manual. Retrieved from http://web.princeton.edu/sites/ehs/labsafetymanual/.

Protection of Human Subjects, 45 C.F.R. § 46 (2009).

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607–620. doi:10.1037/0021-9010.80.5.607.

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*, 445–493. doi:10.1111/j.1744-6570.2012.01250.x.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251. doi:10.1037/0033-2909.87.2.245.

Texas A&M University (2009). Laboratory safety manual. Retrieved from https://ehsd.tamu.edu/Pages/LabSafety.aspx.

Texas A&M University. (2012). EHS laboratory safety inspection checklist. Retrieved from https://ehsd.tamu.edu/Pages/LabSafety.aspx.

Texas Tech University (n.d.). Laboratory safety manual. Retrieved from http://www.depts.ttu.edu/ece/.

The Ohio State University (2014). Institution laboratory biosafety manual. Retrieved from http://ehs.osu.edu/manuals.aspx.

Tourangeau, R. (1984). Cognitive science and survey methods. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey design: Building a bridge between disciplines* (pp. 73–100). Washington, DC: National Academy Press.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*, 299–314. doi:10.1037/0033-2909.103.3.299.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

University of Texas at Austin (2013). Laboratory safety manual. Retrieved from http://www.utexas.edu/safety/ehs/.

US Bureau of Labor Statistics. (2015a). *Census of fatal occupational injuries*. Washington, DC: Government Printing Office **Retrieved from https://www.bls.gov/news.release/cfoi.toc.htm**.

US Bureau of Labor Statistics. (2015b). *Nonfatal occupational injuries and illnesses involving days away from work*. Washington, DC: Government Printing Office **Retrieved from https://www.bls.gov/news.release/pdf/osh2.pdf**.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70. doi:10.1177/109442810031002.

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*, 761–769. doi:10.1016/S0001-4575(99)00129-3.

West Virginia University (2012). Department of physics: Laboratory safety plan. Retrieved from http://physics.wvu.edu/.

Wills, A. R., Watson, B., & Biggs, H. C. (2006). Comparing safety climate factors as predictors of work-related driving behavior. *Journal of Safety Research, 37*, 375–383. doi:10.1016/j.jsr.2006.05.008.

Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology, 65*, 96–102. doi:10.1037/0021-9010.65.1.96.

Zohar, D. (2003). Safety climate: Conceptual and measurement issues. In J. C. Quick & L. E. Tetrick (Eds.), *Handbook of occupational health psychology* (pp. 123–142). Washington, DC: American Psychological Association.

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis & Prevention, 42*, 1517–1522. doi:10.1016/j.aap.2009.12.019.