Pergamon

www.elsevier.com/locate/jsr

Journal of Safety Research 35 (2004) 497–512

www.nsc.org

Exploratory analysis of the safety climate and safety behavior relationship

M.D. Cooper*, R.A. Phillips

Received 1 June 2004; accepted 11 August 2004

N.B. This work is dedicated to the memory of Jonathon 'Flip' Fillippo, an IU Safety Student who died in Jan 2004 while helping others.

Abstract

Problem: Safety climate refers to the degree to which employees believe true priority is given to organizational safety performance, and its measurement is thought to provide an "early warning" of potential safety system failure(s). However, researchers have struggled over the last 25 years to find empirical evidence to demonstrate actual links between safety climate and safety performance. *Method*: A safety climate measure was distributed to manufacturing employees at the beginning of a behavioral safety initiative and redistributed one year later. *Results*: Multiple regression analysis demonstrated that perceptions of the importance of safety training were predictive of actual levels of safety behavior. The results also demonstrate that the magnitude of change in perceptual safety climate scores will not necessarily match actual changes (r=0.56, n.s.) in employee's safety behavior. *Discussion*: This study obtained empirical links between safety climate scores and actual safety behavior. Confirming and contradicting findings within the extant safety climate literature, the results strongly suggest that the hypothesized climate-behavior-accident path is not as clear cut as commonly assumed. *Summary*: A statistical link between safety climate perceptions and safety behavior will be obtained when sufficient behavioral data is collected. *Impact on Industry*: The study further supports the use of safety climate measures as useful diagnostic tools in ascertaining employee's perceptions of the way that safety is being operationalized.

Keywords: Safety performance; Safety climate perceptions; Safety behavior; Organizational accidents; Applied psychology

1. Introduction

Safety climate (Zohar, 1980) is a term used to describe shared employee perceptions of how safety management is being operationalized in the workplace, at a particular moment in time (Byrom & Corbridge, 1997). These descriptive perceptions provide an indication of the "(true) priority of safety" (Zohar, 2000) in an organization with regard to other priorities such as production or quality. Safety climate is considered to be a sub-component of the "safety culture" construct (International Atomic Energy Agency [IAEA], 1988) by some (Cooper, 2000; Glendon & Stanton, 2000; Neal, Griffin, & Hart, 2000; Silva, Lima, & Baptista, 2004; Zohar, 2000) or a reflection of actual safety culture by others (Arboleda, Morrow, Crum, & Shelley, 2003; Cabrera & Isla, 1998; Cox & Flin, 1998; Fuller & Vassie, 2001; Guldenmund, 2000; Lee & Harrison, 2000; O'Toole, 2002; Vredenburgh, 2002; Williamson, Feyer, Cairns, & Biancotti, 1997).

Over the last 25 years, safety climate research has taken four directions: (a) Designing psychometric measurement instruments and ascertaining their underlying factor structures (e.g., Brown & Holmes, 1986; Coyle, Sleeman, & Adams, 1995; Dedobbeleer & Beland, 1991; Garavan & OBrien, 2001; Zohar, 1980); (b) Developing and testing theoretical models of safety climate to ascertain determinants of safety behavior and accidents (e.g., Cheyne, Cox, Oliver, & Tomas, 1998; Neal et al., 2000; Prussia, Brown, & Willis, 2003; Thompson, Hilton, & Witt, 1998); (c) Examining the relationship between safety climate perceptions and actual

* Corresponding author. Dept. of Applied Health Sciences, HPER 116E, Indiana University, East 7th Street, Bloomington, IN 47405, United States. Tel.: +1 812 856 4887.

E-mail address: domcoope@indiana.edu (M.D. Cooper).

0022-4375/$ - see front matter © 2004 National Safety Council and Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsr.2004.08.004

safety performance (Glendon & Litherland, 2000; Zohar, 2000); and (d) exploring the links between safety climate and organizational climate (Neal et al., 2000; Silva et al., 2004).

1.1. Factor structures

Factor analysis is a statistical technique used to identify a relatively small number of non-observable, underlying factors that characterize underlying "constructs" (e.g., management attitudes to safety). In safety climate research, these "factors" are used to represent relationships among many sets of inter-related perceptual questions about safety. The identified "factors" simplify interpretation of these relationships by reducing the observed correlations into as few "constructs" as possible. As yet there is no universal consensus about a key set of underlying factors for the concept of safety climate, or even if one exists (Coyle et al., 1995). It has been suggested that analogous to the personality literature (Barrick & Mount, 1991) there is a Big 5 safety climate structure (Flin, Mearns, O'Connor, & Bryden, 2000). An excellent review (Guldenmund, 2000) indicates the complexity of safety climate as a psychological construct and exposes the Big 5 concept as somewhat premature. Many researchers point to the differences between the factor structure in Zohar's (1980) study and those of Brown and Holmes (1986) and Dedobbeleer and Beland (1991) that used the same instrument (or variants) on different populations in different industries and countries to justify why their factor structure differs from other instruments. However, reported differences in the key underlying factor structures may simply reflect methodological differences in question generation (e.g., focus group exercises, literature reviews), sample populations (within or between companies) across industries, and the labeling of constructs according to the theoretical model driving the research (Guldenmund, 2000). In most instances, the purpose of measuring safety climate is to provide opportunities for enquiry or change (Carroll, 1998) so as to improve safety performance in the measured organization. This means that industrial organizations are the major stakeholders of safety climate research. As such, it is very important that a safety climate factor should only be viewed as key *if it predicts actual, or ongoing, safety performance in organizations.*

Three studies have attempted to validate their factor structures by replication (Rummel, 1970) at two points in time with different (Coyle et al., 1995) or the same (Glendon & Stanton, 2000; Thompson et al., 1998) sample populations. Coyle et al. (1995) obtained a different factor structure, whereas both Glendon and Stanton (2000) and Thompson et al. (1998) obtained similar structures. Such results imply that obtained safety climate factor structures are specific to particular industries and/or sample populations (McDonald & Ryan, 1992) or that different instruments measure distinctly different safety climate concepts (Glendon & Litherland, 2000).

1.2. Validation

Validity is a quality standard for evaluating safety climate and other psychometric measures that refers to their accuracy and appropriateness for predicting or drawing inferences about certain criteria (e.g., safety performance). Relying solely on discriminant validity (e.g., statistical differences in scale scores by demographic variables such as age) most studies have made no attempt to assess the concurrent or predictive validity of their instrument, or identified factors, against independent variables such as accident rate or other measures of safety performance (Brown & Holmes, 1986; Carroll, 1998; Cheyne et al., 1998; Cox & Cheyne, 2000; Cox & Cox, 1991; Dedobbeleer & Beland, 1991; Donald & Canter, 1994; Fuller & Vassie, 2001; Hayes, Perander, Smecko, & Trask, 1998; Mearns, Flin, Gordon, & Fleming, 1998; Neal et al., 2000; Prussia et al., 2003; Rundmo, 1992, 1994; Williamson et al., 1997). The universality of discriminant validity reported across all published studies suggests that sub-group differences within the same organization are a given. This accords with the purpose of safety climate measurement: to identify and explore such differences so as to implement the appropriate remedial interventions (Budworth, 1997). Psychometric [safety climate] instruments are deliberately designed to discriminate between people on various demographic dimensions (Cook, Hepworth, Wall, & Warr, 1993). Thus, any between sub-group differences merely inform about the degree to which the measure has reached its initial design goals. They do not inform about the ability of the measure to assess or predict actual ongoing safety performance. Moreover, correlating demographic data collected at the same time as responses to safety climate questions are collected is not concurrent validation (Bausell, 1986, p216).

Some researchers have attempted to assess concurrent validity (i.e., safety performance at the time of distribution) or predictive validity (i.e., forecast future safety performance) by correlating the scale or factor scores against actual accident rates (e.g., Lee & Harrison, 2000; Mearns, Whitaker, & Flin, 2003; Niskanen, 1994; O'Toole, 2002; Silva et al., 2004; Varonen & Mattila, 2000; Vredenburgh, 2002; Zohar, 2000), expert ratings (Arboleda et al., 2003; Diaz & Cabrera, 1997; Zohar, 1980), human error analysis (Glendon & Stanton, 2000), ratings of behavioral compliance (Garavan & OBrien, 2001), and actual safety behavior (Glendon & Litherland, 2000). With the exception of Zohar (2000) who found a predictive relationship with "micro-accidents" five months after distribution, no safety climate instrument has yet been found to predict actual safety behavior or ongoing levels of safety performance.

1.3. Safety climate modelling

Attempts to delineate the underlying safety climate constructs and their relationships with self-report indices of safety activity have been undertaken using *a priori* Structural Equation Modelling (SEM). Safety activities include subjective appraisals of the physical work environment and workplace hazards (Brown, Willis, & Prussia, 2000; Cheyne et al., 1998), managerial assessments of employee's safety compliance (Prussia et al., 2003), safety hazards and self-reported compliance (Neal et al., 2000; Thompson et al., 1998), and safety participation (Neal et al.). In the same way that differences are reported in factor structures, vast differences are found in theoretical models derived from this process. Importantly, in all of these studies the path correlations between safety climate and the self-report safety activities show the degree of association between constructs to be moderate at best. Given that correlations between two perceptual constructs tend to be inflated (Miller & Monge, 1986) these modelling results may even be over-estimates of actual relationships. Nonetheless, some (e.g., Glendon & Litherland, 2000) argue that the utility of SEM resides in the revelation that safety climate exerts an indirect effect on safety behavior, which is mediated by further variables such as transformational leadership (Barling, Loughlin, & Kelloway, 2002), the work context (Hofmann & Stetzer, 1996), and production pressures (Brown et al., 2000). The validity of the various SEM models is difficult to ascertain as all are based on self-report instruments and none have used independent variables to verify any relationships obtained. However, all the SEM studies report that the relationship between safety climate and safety activity (i.e., behavior) is mediated by other variables. Overwhelmingly, this body of evidence suggests that there is no direct link between perceptual safety climate constructs and actual safety behavior.

1.4. Safety performance

Measurement of safety performance is notoriously problematic as measures such as accident rates and compensation costs tend to be reactive (after the event) and relatively infrequent. This focus on safety results (Cohen, 2002) often means that the success of safety is measured by lower levels of system failure. Many modern approaches (Strickoff, 2000) advocate the use of proactive measures (e.g., safety climate, hazard identification and /or observed percent safe behavior) that focus on current safety activities to ascertain system success rather than system failure. In combination both approaches can help organizations to ascertain the effects of their safety programs.

Derived from behavioral safety, the observed percent safe score is thought to be one of the most useful indicators of current safety performance (Reber, Wallin, & Duhon, 1989). Based on randomized behavioral sampling, employee observers record the number of safe and unsafe behaviors performed by their peers, against predetermined checklists of safety related behaviors derived from accident/incident reports. The observation results are used to compute a percent safe score, which is used in many ways (e.g., set improvement targets) but is primarily intended to provide ongoing feedback (Cooper, Phillips, Sutherland, & Makin, 1994) so that people can adjust their performance accordingly. Reviews of behavioral safety studies have demonstrated dramatic improvements in safety performance (Grindle, Dickinson, & Boettcher, 2000; McAfee & Winn, 1989; Sulzer-Azaroff, Harris, & Blake-McCann, 1994) in terms of reductions in accidents, workers compensation costs, and insurance premiums. To date, no published study has established a clear direct link between measures of safety climate and actual safety behavior. The reported relationships between safety climate and safety behavior have largely been inferred from structural equation models based on a variety of self-report instruments. The notable exception to this trend is Glendon and Litherland's (2000) attempt to measure both safety climate perceptions and actual safety behaviors in road construction. Contrary to expectations, but in accordance with structural equation models, this study failed to establish a direct relationship between the two. The authors speculated that the information obtained from the two forms of measurement is so independent that safety climate and safety behavior exist independently under a super-ordinate safety construct (Culture?). However, the authors also postulated that the number of observations conducted over the course of one day in five-minute periods violated recommendations (Tarrants, 1980) for this type of measurement. This suggests two competing explanations: (a) there is no direct link between safety climate and safety behavior or (b) that a relationship between safety climate scale scores will be found if behavioral measurements are taken over longer periods of time.

1.5. The present study

The present study is an extension of behavioral safety (Cooper et al., 1994) and safety climate work, utilizing a modified version of Zohar's (1980) safety climate instrument (Cooper & Phillips, 1994) carried out in a manufacturing facility. Within the same organization, at the same time that a behavioral safety intervention was conducted, a safety climate survey was completed. The intervention was designed and implemented as a continuous process that would eventually involve all plant personnel in safety observations for a period of three to five months each, in order to improve levels of safe behavior. Twelve months after the first safety climate survey a second survey was conducted with the original survey instrument. Archival behavioral safety data made available to the first author by the study organization presented an opportunity to test various hypotheses pertaining to the current status of safety climate research. Based on the current safety climate literature, the following hypotheses were formulated.

1. When used to survey the same sample population, at different points in time, a similar factor structure will be obtained from the same safety climate instrument.

2. Differences in perceptions will be demonstrated across a variety of demographic variables such as self-reported accident involvement, age, job experience, and functional departments, for both pre and post distributions of the safety climate measure.
3. No direct relationship will be obtained between safety climate scale scores and safety behavior.
4. Measured changes in safety behavior will not be reflected in similar changes in pre and post test safety climate scores.

2. Method

2.1. Sample

The study population was the plant personnel (n=540) of a packaging production plant. The questionnaire, along with return envelopes addressed to the researchers, was initially distributed through the internal mail system to all members of the organization. At the same time a behavioral safety initiative was being implemented across the organization. The response rate for this distribution was 69% (n=374). The average age of all respondents was 45.3 years (SD=10.22) with a range from 18–63 years. Average experience in their current job was 11.72 years (SD=9.96). Exactly the same methodology was employed for the second distribution. The response rate for this distribution was somewhat lower at 35% (n=187). The average age of all respondents was 42.21 years (SD=11.34) with a range from 19–64 years. Average experience in their current job was 12.8 years (SD=10.69).

2.2. Safety climate measure

A 50-item measure was developed specifically for the plant, originally based upon the work of Zohar (1980). Of the original 40 items that comprised Zohar's questionnaire, 28 were used in the current measure, either directly or with minor changes to the wording. Some of Zohar's original managerial attitude items were more concerned with managerial actions and behavior as opposed to attitudes, a view supported by Brown and Holmes (1986) and Dedobbeleer and Beland (1991). These led to major changes to the managerial attitude scale and the development of a new management actions scale. These actions were concerned with the extent to which management encouraged workers to behave safely or unsafely. The extent of item change is shown in Table 1. The final measure comprised 50 items, scored on a 5-point Likert type scale; on a continuum from highly disagree to highly agree. Twelve of the items were negatively worded and as such were reverse scored. In addition, six demographic variables concerned with department of respondent, current job title and years of experience in job, whether or not the respondent had been involved in an accident, and if so how many and what type, sex of respondent, and age. A cover letter was also attached assuring all respondents of their anonymity.

2.3. Observed percent safe

Previously reported in Cooper et al. (1994) a behavioral safety initiative was implemented across the business. In brief, this entailed developing separate behavioral observation checklists for nine main departments in conjunction with the workforce (see Fig. 1 for an example). An employee from each workgroup within each of the departments was trained to be an observer (n=48). The observer monitored everyone within their respective work areas for 10–20 minutes every day they were at work. A behavior was recorded as safe *only* if everyone in a workgroup was performing that behavior safely. The observed percent safe scores (# of safes / # of safes + # of unsafe's x 100) recorded for the first 4–6 week shift rotations were used to establish "baselines" by which participative safety improvement goals could be set and future performance compared (n=20 weeks). Results showed statistically significant improvements in safety performance and decreases in accident rates. The initiative continued "in-house" for a further 3–4 years until the business was acquired by a third party but the available results have not been previously reported. For each and every subsequent 20-week period, employees adapted the observation checklists for their work area. Those behaviors recorded as 100% safe for an extended period of time (n=10 weeks) were removed and replaced with others identified during the observations or obtained from accident/near-miss records. Thus, the behavioral checklists used at the time of the second distribution of the safety climate questionnaire were somewhat different to those in the first distribution. New observers were also introduced at the same time. In effect, the initiative was constantly being refreshed in an attempt

Table 1
Number of changes to Zohars (1980) 40 item safety climate measure

Factor	Total # of Zohar (1980) Items	# of Omitted Items	# of New Items	Total # of Study Items
Management attitudes towards safety	9	3	6	12
Management actions towards safety *(New Scale)*	0		4	4
Perceived level of risk at workplace	5	1	2	6
Effects of required work pace on safety	3	3	4	4
Importance of safety training	6	1	2	7
Effects of safe conduct on social status (2) & promotion	9	3	2	8
Status of safety officer & safety committee	8	1	2	9
Totals	***40***	***12***	***22***	***50***

Department: Finishing/sheeting.

		Safe	Unsafe	Not Seen
1]	Goggles must be worn when using nail gun.	-----	------	---------
2]	Gloves must be worn when handling pallets and cases.	-----	------	---------
3]	Electric knife must be used to cut film off the drums, whenever possible.	-----	------	---------
	a] A manual knife may be used to start or finish the cut.			
4]	Nail guns only to be used in safe area behind protective barriers.	-----	------	---------
5]	Loose strapping on pallets must be removed.	-----	------	---------
6]	Protruding nails in pallets etc. must be bent over safely or removed.	-----	------	---------
7]	No jewelry should be worn (rings, watches, necklaces, etc).	-----	-------	---------
8]	Safety shoes must be worn.	-----	------	--------
9]	Drivers of small forklift trucks must always be at the front of vehicle, when travelling in any direction.	-----	-------	---------
10]	Reach trucks not to be driven with forks raised.	-----	-------	---------
	with forks extended.	-----	-------	---------
11]	Horns must be sounded on 'blind' corners	-----	-------	---------

Fig. 1. Example behavioral safety checklist.

to embed the process into the "normal" way of doing things so that it would not be viewed as "flavor of the month." Company management provided all the necessary training for observers and any resources required.

3. Results

3.1. Safety climate

Utilizing the Statistical package for the Social Sciences (SPSS), the data were analyzed to evaluate the safety climate instruments factorial structure for each distribution. The method of factor extraction was principal components, rotated according to a varimax solution when two or more factors emerged. In addition, a "second order" factor analysis of the scales was conducted. The overall reliability of the measure for both distributions was also assessed using Cronbach's Alpha.

3.1.1. Factor analytic results

The analyses shows that in both distributions Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was 0.82 indicating that the data were appropriate for this analysis (Kaiser, 1974). Bartlett's Test of Sphericity was significant for both the pre [χ^2=632.67, p<.0000] and post [χ^2=1016.73, p<.0000] test, indicating that correlations exist among some of the safety climate scales. To enhance interpretability, only factor loadings greater than 0.5 were selected (Norusis, 1985). In both distributions the same two second order factors were obtained, which together accounted for 69.9% and 68.2% of the variance, respectively, (see Table 2) supporting hypothesis 1, that a similar factor structure will be obtained from the same safety climate instrument when used to survey the same sample population, at different points in time. The first factor comprised: (a) perceived levels of risk in the workplace; (b) perceived management attitudes toward safety; (c) the effects of the required work pace on safety; (d) management actions toward safety; and (e) the perceived importance of safety training. The second factor comprises (a) the perceived importance of safety training; (b) perceived effects of safe conduct on social status and promotion; and (c) the status of the safety officer and safety committee. On the basis of these findings it could be argued that workers make a major distinction between factors that directly affect their perceptions about the way safety is operationalized and those that affect their perceptions indirectly. Given the factor analytic results were the same for both distributions, the results suggest that the underlying constructs of this safety climate measure are robust thereby psychometrically validating the measure as recommended by Rummel (1970).

3.1.2. Internal reliability

The reliability of a measure refers to its internal consistency. Consistency is assessed by the manner in which all employees respond in similar ways to similar questions that measure a particular construct (e.g., Manage-

Table 2
Second order factor analyses of the safety climate measure

	Pre-test		Post-test	
	Factor 1	Factor 2	Factor 1	Factor 2
Perceived level of risk	0.83		0.82	
Management attitudes toward safety	0.81		0.84	
Effects of work pace	0.79		0.76	
Management actions toward safety	0.74		0.79	
Importance of safety training	0.57	0.53	0.56	0.51
Social status & promotion		0.87		0.87
Safety officer & committee		0.72		0.62
Eigen Value	3.79	1.11	3.77	1.00
% of Variance explained	54.1	15.8	53.9	14.3
Sum of variance Explained	69.9		68.2	

ment's commitment). Cronbach's coefficient alpha (Cronbach, 1951) is used when questions are rated on interval scales such as five point Likert scales and represents the reliability coefficient that would have been obtained from all possible combinations of dividing the questions into two sets (split-halves). For example, dividing the questionnaire into all odd and even numbered questions and correlating them. Although a measure cannot be too reliable, as a rule, reliability coefficients around 0.7 and above are professionally acceptable (Muchinsky, 2004). Utilizing Cronbach's Alpha, the overall internal reliability of the safety climate measure for the first distribution (pre-test) was 0.925 and 0.933 for the second (post-test). This suggests that the measure consistently measures what it is supposed to measure. The similar reliabilities obtained for the two different distributions suggest it also possesses temporal stability over time. The reliability of each individual scale for both distributions is reported in Table 3. The reliability coefficients for the scales concerned with management attitudes and perceived levels of risk are very good. With Cronbach Alpha coefficients greater than 0.70, the scales of effects of work pace and management actions are also acceptable in psychometric terms. However, the scales concerned with safety training and the status of safety officers and committee's, although acceptable as research instruments, appear to require further development. The scale items concerned with the social status of safety and promotion appear to need considerable rework. For the second distribution (post-test), the reliability of each scale is of a similar magnitude to that of the first distribution although the data appears to show a trend of increasing reliability, which is most likely related to less overall variance amongst the item scores in the second distribution. Tests of significance were conducted on the differences between the reliability coefficients on all the scales. The two scales of perceived levels of risk and management actions were significantly different at the .05 level. No significant differences were found for the remaining scales.

3.2. Discriminant validity

To test hypothesis 2 with regard to sub-group differences and the instruments discriminant validity, the safety climate data were analyzed with an independent groups one-way analysis of variance (ANOVA) procedure (see Table 4). ANOVA's are designed to test differences between several groups of mean average scores and are based on the ratio of between group variability and within group variability. A significant F statistic signals only that the group means are unequal (i.e., different), it does not pinpoint where the differences are. As it requires large mean differences, the 'Scheffe' multiple comparison test was used to identify which groups were significantly different from each other. The ANOVA procedures focused upon whether or not statistically significant differences in safety climate existed between sub-groups such as age groups, employees with different lengths of job experience, those who have or have not been involved in accidents (Brown & Holmes, 1986), and the various functional departments of the same organization (Drexler, 1977; Joyce & Slocum, 1984; Newman, 1975; Schneider & Bowen, 1985).

Table 3
Cronbach's Alpha for each safety climate dimension

	Pre-test	Post-test
Management Attitudes toward Safety	0.90	0.91
Perceived level of Risk	0.86	0.90*
Effects of work pace	0.73	0.72
Management Actions toward safety	0.72	0.81*
Safety Officer & Committee	0.65	0.73
Importance of safety training	0.60	0.62
Social status & Promotion	0.50	0.59
Overall Reliability Co-efficient	*0.925*	0.933

* Statistically significant differences between Cronbach Alpha's ($p<.05$).

3.2.1. Differences by age

Four age groups (16–34 yrs; 35–44 yrs; 45–54 yrs; 55–65 yrs) were treated as different levels of the factor. No significant differences were found in the pre-test distribution on any of the scales. On the post test distribution, significant differences in perceptions emerged on three scales between the youngest group (who had the lower perceptions) and the three older groups. These were management attitudes, management's actions, and safety training. Overall, this particular analysis does not support hypothesis 2. However, the results do suggest that behavioral safety interventions exert positive effects upon the perceptions of older workers yet have little impact upon those in the youngest age group.

3.2.2. Differences by years of job experience

Four categories of years job experience (1–5 yrs; 6–15 yrs; 16–25 yrs; 26–45 yrs) were entered as different levels of the factor. Significant differences emerged in the pre-test distribution, on four scales (management attitudes and actions, levels of risk, and safety training). No significant differences emerged on the work pace, safety officer/committee, or social status scales. However, on the post test distribution significant differences in perceptions emerged on all the scales except social status. Overall, this analysis partially supports Hypothesis 2.

3.2.3. Differences by accident involvement

Four accident involvement groups (no accident personnel, personnel involved in minor accidents, major accidents, or both types of accident) were entered as different levels of the factor. Significant differences emerged in the pre-test distribution, on four scales (management attitudes and actions, levels of risk, and work pace). No significant differences emerged on the safety training, safety officer/committee, or social status scales. As shown in Table 4,

Table 4
Demographic differences by safety climate scales

Dimension	Pre-test			Post-test		
	D.F.	F Ratio	P<	D.F.	F Ratio	P<
Age						
Management Attitudes	3, 260	0.87	n.s.	3, 179	4.07	.01
Levels of Risk	3, 264	0.55	n.s	3, 179	0.33	n.s.
Management Actions	3, 261	2.15	n.s.	3, 179	5.83	.01
Workpace	3, 268	1.93	n.s.	3, 179	1.60	n.s.
Safety Training	3, 262	1.86	n.s.	3, 179	2.66	.05
Safety officer/committee	3, 254	1.64	n.s.	3, 179	1.72	n.s.
Social status/promotion	3, 259	0.36	n.s.	3, 179	2.05	n.s.
Years Job Experience						
Management Attitudes	3, 236	3.46	.01	3, 180	5.37	.01
Levels of Risk	3, 243	8.78	.01	3, 180	6.09	.01
Management Actions	3, 238	3.93	.01	3, 180	2.97	.05
Workpace	3, 246	1.83	n.s.	3, 180	6.71	.01
Safety Training	3, 239	2.69	.05	3, 180	3.85	.01
Safety officer/committee	3, 231	.58	n.s.	3, 180	2.85	.05
Social status/promotion	3, 241	1.60	n.s.	3, 180	1.32	n.s.
Accident Involvement						
Management Attitudes	3, 264	2.91	.05	3, 180	1.47	n.s.
Levels of Risk	3, 269	10.72	.01	3, 180	8.51	.01
Management Actions	3, 266	2.55	.05	3, 180	1.15	n.s.
Workpace	3, 274	7.58	.01	3, 180	4.14	.01
Safety Training	3, 267	1.74	n.s.	3, 180	3.50	.05
Safety officer/committee	3, 257	0.01	n.s.	3, 180	0.67	n.s.
Social status/promotion	3, 263	0.27	n.s.	3, 180	0.86	n.s.
Functional Departments.						
Management Attitudes	6, 349	9.42	.01	6, 179	7.36	.01
Levels of Risk	6, 356	34.49	.01	6, 179	16.40	.01
Management Actions	6, 351	5.66	.01	6, 179	3.62	.01
Workpace	6, 361	8.68	.01	6, 179	3.46	.01
Safety Training	6, 352	5.57	.01	6, 179	2.83	.01
Safety officer/committee	6, 341	2.21	.04	6, 179	1.28	n.s.
Social status/promotion	6 343	.38	n.s.	6, 179	1.30	n.s.

N.B. Sample sizes differ due to non-reporting by respondents.

however, significant differences in perceptions emerged on only three scales (levels of risk, work pace, and safety training) on the post test distribution. This suggests that management attitudes and actions were seen as more positive by the accident involvement groups as a result of introducing the behavioral safety intervention, but little changed in regard to perceived risk levels or work pace. Again, this analysis provides partial support for hypothesis 2.

3.2.4. Differences by functional department

The various departments were treated as levels of the factor. Analyses were conducted on both pre and post test measures separately. The results of these were highly significant on all the scales, except for the scales related to social status on promotion and the status of safety personnel. These results support hypothesis 2 and suggest the existence of different safety climates within different departments of this organization.

The results of the various groupings (Department, Age, Job Experience, and Accident Involvement) demonstrate that the safety climate instrument used in this study possesses discriminate validity although hypothesis 2 was only partially supported. Interestingly, there appears to be more consistent discriminant validity when organizational demographics (i.e., Department) are used as opposed to personal demographics (Age, Accident Involvement, and Years Job Experience). This is likely related to the different risk levels associated with various tasks and activities that take place in each separate department.

3.3. Links between safety climate and safety behavior

The safety climate survey was distributed, and responded to, over two four week periods 12 months apart. The observed percent safe data obtained during these 4-week periods is used as the basis for all comparisons between the behavioral and climate scores. These analyses focus on safety behavior first followed by those for safety climate. Some departments had too few respondents to meaningfully analyze the results, while some indicated their work area rather than department. In such instances respondents were grouped according to the main department in which they worked prior to any analysis being undertaken.

3.3.1. Levels of analysis

The level of analysis and aggregation procedures is a continuing concern among climate researchers: climate can be investigated at many organizational levels of analysis (Rousseau, 1985). Although survey data is collected from individuals, it can also be aggregated to provide limited information about the actual activities of focal units (e.g., work groups, departments, divisions, or an organization as whole). In accordance with Zohar (2000) the conditions determining the appropriate level of analysis require: (a) within-group homogeneity (i.e., the degree to which all group members equally share perceptions about the various facets of safety climate) and (b) between-group variance (the degree to which safety climates differ significantly between one sub-unit and another in the same organization). In this study these conditions have been met as demonstrated by the F Ratio's obtained for the significant differences found between departments. This indicates that department is the appropriate level of analysis to use to attempt to establish any links between climate perceptions and actual behavior.

3.3.2. Behavioral safety results

Upon completion of the first observation period (phase 1) the company concerned changed the behaviors on the departmental observation checklists and trained further observers. Under guidance of the first author they followed exactly the same procedures as those specified in phase 1 (Cooper et al., 1994). They continued this process for all subsequent phases, independently of the researchers. The global results of these efforts for the whole site are shown in Fig. 2 from weeks 1–76. Phase 2 results showed continued improvement in the percent safe scores. Phase 3 and 4 results show somewhat lower observed percent safe scores in general, when the company implemented solely by themselves, but still improving from each phase baseline (Phase 4 shows only the first 10 weeks data, as this was all that was available to the researchers). In terms of reported accidents (minor and major), there was a slight increase in phase 2 (from 63–69), but reductions in phase 3 (n=57) and again in phase 4 (n=46 for the whole phase). Ironically, lower percent safe scores were being recorded as the number of accidents per phase was falling; although this may be due to impact time lags (Duff, Robertson, Phillips, & Cooper, 1994) resulting from the previous phases. As previously reported however, this further illustrates that there is not necessarily a direct link between levels of observed percent safe and recorded accident rates (Cooper et al., 1994).

3.3.3. Data transformation

To ensure 100% correspondence between the safety climate and observed percent safe scores, some of the behavioral percent safe scores from various workgroups were amalgamated (Cooper et al., 1994). Specifically, personnel from "Slitting, Sheeting and Reelwrap" were grouped as the Finishing department; Production office and Laboratory were grouped together under Production Offices; Non-production offices included Information Technol-

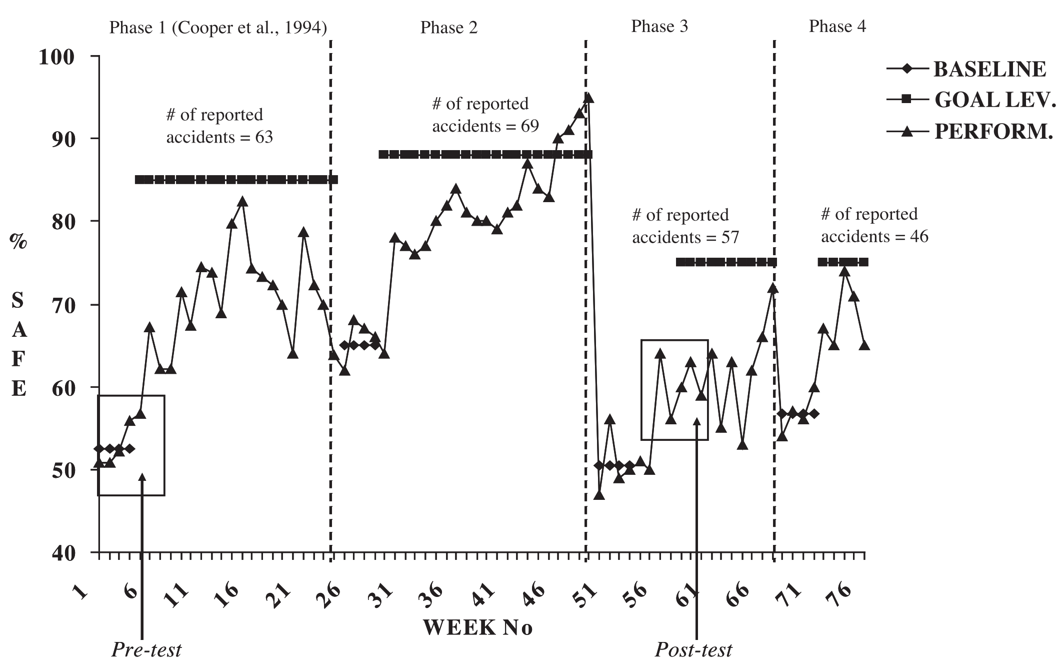

Fig. 2. Global observed percent safe results over four phases.

Table 5
Observed percent safe scores at time of safety climate distributions

Department	1	2	3	4	Pre-test		1	2	3	4	Post Test	
					Mean Ave	*SD*					*Mean Ave*	*SD*
Casting	22	28	28	28	*27*	*3.0*	54.5	48	49.5	55.5	*51.88*	*3.68*
Coating	67	58	74	71	*69*	*6.95*	47	38	66	57	*52*	*12.14*
Engineers	40	45	41	61	*46*	*9.74*	73	C	C	C	*93.25*	*13.5*
Finishing	63	62	64.5	57.6	*61.8*	*2.97*	84	81.67	81	80	*81.67*	*1.70*
Non Prod Office	73.6	72.6	66.3	79.6	*73.6*	*5.45*	84	C	94	C	*94.5*	*7.55*
Production Office/Lab	65.5	89	79.5	76.5	*78*	*9.68*	92	C	93	C	*96.25*	*4.35*

KEY: C=100%.

ogy, Sales & Accounts, CSD & Human Resources. Viscose and Day gang were excluded from these analyses as no completed surveys were received from this group in the second distribution of the safety climate measure. Table 5 shows the amalgamated observed percent safe scores by department during the pre and post test safety climate distributions. The departmental average scores were used as the basis for all subsequent analyses.

3.3.4. Concurrent validity

To test the hypothesis that there is no direct link between safety climate constructs and safety behavior, the factor scores and individual scale scores for each department were correlated with the equivalent observed percent safe scores, obtained during the safety climate distribution periods. This analysis provides an estimate of the instruments concurrent validity (i.e., the relationship between climate and behavior scores at the same point in time). Validity coefficients over 0.5 are considered rare (Muchinsky, 2004).

Shown in Table 6, significant correlations were obtained between Factor 1 (r=.88, p<.05) and Factor 2 (r=.99, p<.01) and the behavioral safety scores for the first distribution. Only Factor 1 (r=.90, p<.05) obtained statistical significance during distribution two. The shared variance (r^2) for the safety behavior scores and the two safety climate factors in distribution one of 77% and 97% respectively, and 81% for Factor 1 in distribution two strongly suggest that employee's perceptions of safety climate (as measured) were consistently associated with actual ongoing safety performance at the time of each distribution. Contrary to SEM studies, this indicates that there *is* a direct link between safety climate and safety behavior. Factor 1 appears to be consistently associated with ongoing performance, whereas Factor 2 appears to reflect actual performance at the starting point of a behavioral safety intervention.

To ascertain the degree of association between the individual safety climate scales within each factor and actual ongoing safety performance, correlations were again conducted for each distribution. Presented in Table 7, the correlations ranged from 0.71 to 0.91 in the first distribution. However, only the scales related to management actions, the importance of safety training, and the status of safety personnel were significant (p<.05). The shared variance (r^2) for both safety behavior scores and these three safety climate scales range from 66%–83%. For the second distribution, with the exception of the scales concerned with the status of safety personnel (r=0.31) and the effects of safe conduct on promotion (r=0.22), concurrent validity coefficients ranged from 0.70–0.94. Significant correlations were obtained for management attitudes, management actions, and perceptions of risk scales (p<0.5), and the importance of safety training (p<0.1). The shared variance (r^2) for these latter four scales ranged from 67%–88%.

The scales for management actions and safety training were the only two that were significantly associated with actual safety performance for both distributions, suggesting that these two scales in combination might explain the high validity coefficients obtained for Factor 1 in the above analyses. The safety training scale appears to most closely match actual levels of performance in both distributions (0.91 and 0.94), and might also indicate that behavioral safety initiatives are perceived as a form of safety training, rather than a culture change or behavioral change 'initiative' *per se*.

3.3.5. Predictive validity

Due to the high concurrent validity coefficients, stepwise multiple regression analysis was undertaken to attempt to statistically determine which scales might be considered predictive. It is acknowledged that the small number of the aggregate scores violate normal 'subject-to variable ratio'

Table 6
Correlations between aggregated departmental safety climate factor scores and observed percent safe scores

Factor	Distr	n	Mean	SD	Behavior 1			Behavior 2		
					r	p<	r^2	r	p<	r^2
FACTOR 1	1	6	3.06	.283	.88	.05	0.77			
FACTOR 2	1	6	3.48	.312	.99	.01	0.97			
FACTOR 1	2	6	3.06	.095				.90	.05	.81
FACTOR 2	2	6	3.20	.093				.66	n.s	

Table 7
Correlations between safety climate scale scores and observed percent safe scores (by Department)

Factor	Scale	Pre-test			Post-test		
		r	P<	r^2	r	P<	r^2
1	Management Attitudes	0.75	n.s		0.82	.05	.67
1	Levels of Risk	0.79	.n.s		0.92	.05	.85
1	Management Actions	0.81	.05	.66	0.86	.05	.74
1	Work pace	0.71	n.s		0.70	n.s.	
1 & 2	Safety Training	0.91	.05	.83	0.94	.01	.88
2	Safety officer / committee	0.86	.05	.74	0.31	n.s.	
2	Social status / promotion	0.78	n.s		0.22	n.s.	

assumptions, which may inflate the degree of shared error variance compared to true variance (Bausell, 1986). Moreover, highly correlated independent predictors could indicate multicollinearity (inter-correlated predictors) that can cause severe computational problems. One method of assessing the importance of multicollinearity is to ascertain if the expected residual sum of squares is affected by a large variance inflation factor (VIF) that is greater than 10 (Von Eye & Schuster, 1998). In general, multicollinearity does not prevent the estimation of models of good fit (adjusted R^2), but interpretation of the model becomes largely dependent upon the other predictors that are in the equation. One method of overcoming multicollinearity is to drop those predictors that correlate highly with other predictors. However, this also means that the estimate of the contribution these make to the overall model is prevented. Also the predictors remaining may be severely affected by the absence of the eliminated predictors. As the purpose of this analysis is to explore the data in an attempt to identify those combinations of safety climate scales that are most closely associated with the observed percent safe, all predictor scales were entered into the equation. The observed percent safe scores were entered as the dependent variable with each safety climate scale entered as independent variables.

The results shown in Table 8 indicate that during the first distribution the importance of safety training explained 77.3% of the variance in the observed percent safe scores. The scale concerning safe conduct on promotion explained a further 18.7%. Thus in combination these two scales appear to be predictive of actual safety behavior [$F=58.64$, $p<.01$]. The VIF factor for each of these two scales was 1.301, much lower than the critical value of 10. For the second distribution, only the scale concerning importance of safety training explained any of the variance in observed percent safe scores [$F=30.66$, $p<.01$]. All other scales were dropped as the default SPSS criteria limits had been reached. The VIF value was 1.00, which was again well below the critical value of 10. From the two sets of VIF values it can be concluded that analyses of the present data did not face major multicollinearity problems. Thus, these analyses suggest the link between safety climate and safety behavior appears to arise from perceptions about the importance of safety training in particular. It is noteworthy that despite the significant correlations obtained for both distributions between the perceived management action scale and observed percent safe reported in Table 7, that perceived management actions was not found to be a predictor of actual safety performance (Zohar & Luria, 2003). However, this finding may be confounded by the small "subject-to variable ratio" caused by the use of aggregated departmental datasets.

3.4. Magnitude of change in safety climate and safety behavior scores

To begin to ascertain the extent to which perceptions about safety climate had changed during the first year of the behavioral safety project, t-tests were conducted on the overall scales scores of the pre- and post-test distributions. The results revealed that six scales showed statistically significant changes, with the exception of the status of safety personnel scale, which was non-significant (see Table 9). These results support the validity of the measure as a diagnostic tool in determining current levels of safety climate and also support the notion that behavioral safety interventions do impact upon perceptions of safety climate (Zohar & Luria, 2003).

However, these t-test results do not inform about the extent to which changes in safety climate scores are reflected in changes in the observed percent safe score. The correspondence between the two sets of change scores was determined by the use of treatment effect sizes (i.e., standardized deviation units) used in meta-analysis (Hunter & Schmidt, 1990). The basic aim of a meta-analysis is to aggregate research evidence across, or within studies, to identify the average treatment effect size. To facilitate this

Table 8
Multiple regression results

Predictor scales	Adj. R^2	% diff (+)	Beta	F	P<
First distribution					
Importance of Safety Training	.773	77.3	.687	17.99	.013
Safe Conduct on Promotion	.957	18.7	.451	56.84	.004
Second distribution					
Importance of Safety Training	.856	85.6	.9405	30.66	.0052

Table 9
Global changes in safety climate by scales

Scale	n_1	$\bar{X}_1$	sd_1	n_2	$\bar{X}_2$	sd_2	‘t’	df	p<
Management Attitudes	356	3.51	.72	186	3.73	.70	3.43	540	.01
Levels of Risk	363	2.97	.94	186	3.14	1.02	1.93	547	.05
Management Actions	358	3.43	.67	186	3.60	.69	2.85	542	.01
Work pace	368	3.15	.87	186	3.40	.82	3.30	552	.01
Safety Training	359	2.87	.51	186	3.15	.49	6.05	543	.01
Safety officer/committee	348	3.06	.46	186	3.13	.50	1.32	532	n.s.
Social status/promotion	350	3.18	.41	186	3.27	.38	2.53	534	.01

accumulation of results across or within studies, a common statistic needs to be identified. In the present study Cohen's ‘d’ (Cohen, 1977) was the statistic used. This was derived from differences in mean scores (X_1–X_2) between the pre and post test of the safety climate measure, divided by the pooled standard deviation (Sd_w):

$$Sd_W = \left(\left[(N_1-1)Sd_1^2\right) + \left((N_2-1)Sd_2^2\right)\right]/(N_1+N_2-2)\right)^{1/2}$$

As no pure control group was available and given that the same sample population took part in both the pre and post test, the pooled sample standard deviation was deemed to be the most appropriate (Mullen, 1989). The result of this conversion provides an estimate of the magnitude of difference between two means in relation to the normal spread of scores in the sample. When the magnitude of ‘d’ is greater than 0.1, 0.4, and 0.8, the treatment effect is considered small, medium, and large, respectively (Cohen, 1988). The average effect size for organizational interventions is around 0.44 (Guzzo, Jette, & Katzell, 1985). Credibility intervals are used to identify when moderators are operating and confidence intervals are used to estimate the accuracy of the mean effect size (Whitener, 1990).

The pre and post means for the departmental safety climate scale scores were transformed and analyzed with an adapted version of a meta-analytic computer program contained in Hunter and Schmidt (1990). Shown in Table 10, the global Cohen's ‘d’ was 0.43 (i.e., medium size effect). However, the small credibility interval and the zero included in the confidence interval indicate that moderator analyses are warranted. For the purposes of this study it was assumed that department was the moderator in question. The ‘d’ statistic for each department ranged from 0.02 to 0.57. Other moderators appear to be operating in the Coating, Engineering and Production Office departments, but not for Casting, Finishing and Non-production offices. Although providing useful insights, the purpose of this analyses was to obtain and compare Cohen's ‘d’ for both climate and behavior scores. As such, no further moderator analyses were conducted. The observed percent safe scores obtained “at the time of each safety climate distribution” were transformed to obtain the treatment effect sizes (Cohen's ‘d’) for the site as a whole and for each department. Although a small global effect size (d=0.33) was found, the treatment effect sizes for the departments ranged from −0.09 (no effect) to 0.82 (a large effect), indicating that the magnitude of behavioral changes was considerably different across the departments. These variations in effect size are in accordance with that for the safety climate scores. The small credibility intervals combined with the zero in the con-

Table 10
Treatment effect sizes for department changes

Department	n	k	d	var	Credibility interval	Confidence interval
Safety climate						
Global	464	42	.43	0	.43 to .43	−0.04 to .88
Casting	121	7	.57	.019	.39 to .75	.18 to .92
Coating	56	7	.15	0	.15 to .15	−0.39 to .69
Engineering	71	7	.40	0	.40 to .40	−0.10 to .87
Finishing	93	7	.44	.008	.33 to .56	.01 to .85
Non Production office	67	7	.69	0	.69 to .69	.16 to 1.17
Production office	56	7	.02	0	.02 to .02	−0.53 to .56
Observed percent safe						
Global	*880	6	.33	.096	−.26 to .63	−0.24 to 0.42
Casting	168	1	.51	0	.51 to .51	−0.29 to 0.72
Coating	168	1	−.09	0	−.09 to −.09	−0.30 to 0.13
Engineering	168	1	.21	0	.21 to .21	0.00 to 0.43
Finishing	168	1	.82	0	.82 to .82	−0.60 to 1.03
Non Production office	40	1	.36	0	.36 to .36	−0.15 to 0.57
Production office	168	1	.18	0	.18 to .18	0.03 to 0.40

* n=total number of observations; k=total number of effect sizes.

fidence intervals suggests that moderators such as 'shift' group may be operating, but no further analyses is warranted for the purposes of this investigation.

A non-significant result (r=0.56, n.s.) was obtained when the two sets of departmental treatment effect sizes were correlated. Casting was the only department to show a similar magnitude of change for both behavioral (d=0.51) and climate scores (d=0.57). Difference scores obtained by deducting climate effect sizes from behavioral effect sizes shows that less behavior change was associated with greater change in climate scores in Coating (+0.24), Engineering (+0.19), and Non-production offices (+0.33). Greater behavioral changes was associated with lower climate change scores in Finishing (−0.38) and Production offices (−0.16). These difference scores demonstrate that the magnitude of change in the two variables will not necessarily be in the same direction. The notion that the magnitude of change in behavioral scores will be reflected in similar changes in safety climate scores is not supported either. This undermines the assumption that positive changes in safety climate perceptions will be associated with equally positive changes in behavior or vice versa.

4. Discussion

In accordance with the multiple directions taken by safety climate researchers, this study examined the underlying factor structure of an adapted safety climate measure originally developed by Zohar (1980), and attempted to ascertain the instruments discriminant, concurrent, and predictive validity. This study also explored the relationship between safety climate and safety behavior. Based on previous evidence, four hypotheses were tested.

4.1. Factor analyses

The factor analytic results supported hypotheses 1 in that a similar two-factor structure was obtained for both distributions with similar Eigen values for the factors, similar scale factor loadings, and reliabilities. This replication confirms the psychometric validity of the current two-factor model of safety climate (Rummel, 1970). The first factor comprises facets of safety that employees may use to "directly" assess how safety is operationalized in their organization such as management attitudes and actions, risk levels, speed of work, and the importance of safety training. The second factor comprises of constructs such as the status of safety personnel and the effects of safe conduct on promotion, which may be used to indirectly assess the importance of safety. Interestingly, the importance of safety training was the only scale to load upon both factors. The items comprising this scale were drawn directly from Zohar's (1980) measure that had the largest Eigen value in that study. In the present study, this dimension has the lowest factor loadings. Two possibilities may account for this result. First, it may simply reflect differences between an Israeli and British population. Second, it suggests that the use of stratified sampling across multiple organizations and/or geographical locations (Zohar, 1980) in climate research introduces significant amounts of error variance. This may partially explain why different factor structures emerge from different research groups. Nonetheless, these findings also support Glendon and Litherland's (2000) contention that some safety climate factors are stable across industries and cultures.

4.2. Discriminant validity

In partial support of hypotheses 2, the instrument discriminated between demographic groupings of respondents on some safety climate scales, but not all. During the first distribution, the survey exhibited discriminant validity on some scales for job experience, accident involvement, and department, but not for age. Some scale differences were found for all the demographic groupings during the second distribution. It must be recognized that any non-significant statistical differences may simply reflect the fact that people did not actually differ in their views. Of particular note is the significant differences found between departments in both distributions, suggesting that differences in types of work activity and other localized situational conditions are much more important in climate research than personal demographical variables such as age, job experience, or accident involvement. This finding makes sense as safety climate measures tend to try to capture people's perceptions about how safety is operationalized in their *organization*. They do not tend to measure how the prevailing safety climate affects them as '*individuals*' who have longer work experience, as older or younger workers, or as accident or non-accident victims. Reporting differences in personal demographic variables can be useful and informative for the participating organization but in the first authors experience these tend not to be used in any practical way. Concentrating on organizational demographics such as job function, divisions, and departments is likely to be a more fruitful route to discovering relationships between safety climate and other organizational variables rather than personal variables. Overall, these study results support the notion that some sub-group differences in perceptions of safety climate will almost always be found. However, attempts to validate various safety climate instruments with the sole use of personal demographic sub-group differences does not appear to be particularly informative about how to further improve the content of measures or safety performance *per se*. At the very least, empirical justification for using personal demographics as a validation technique is required if safety climate research is to continue progressing.

4.3. Concurrent validity

Notwithstanding the study limitations, the relationships found between safety climate and safety behavior presented in Table 6 lead to a rejection of hypothesis 3. Contrary to the results of Structural Equation Modeling studies, there *does* appear to be a direct relationship between the safety climate and safety behavior when sufficient data is collected (Glendon & Litherland, 2000). A consistently strong statistical relationship was obtained between observed percent safe and Factor 1 at both time points. As such the scales that comprise Factor 1 may be viewed as key to the measurement of safety climate as they provide an indication of the actual levels of safe behavior exhibited at the time of measurement. Factor 2 was strongly related to levels of safety behavior during the first distribution only. This supports the notion of reciprocal relationships between safety climate and behavior within a hypothesized safety culture model (Cooper, 2000). Moreover, the "direct" and "indirect" interpretations of the factors suggest that employee assessments of social status becomes less important over time as employee-driven improvement initiatives exert their effect.

The results of Factor 1 may be accounted for by the consistent relationships between the scales for management actions and the importance of safety training with behavior at both time points, as presented in Table 7. The results for the effects of management actions on observed percent safe are in accordance with those of Zohar and Luria (2003), which demonstrated relationships between supervisory safety-oriented interactions and significant changes in workers' safety behavior and safety climate scores. However, the multiple regression results presented in Table 8 strongly suggest that perceptions about the importance of safety training in particular are predictive of actual ongoing behavior. This supports climate studies that examined the links between training and self-report measures of safety culture (Arboleda et al., 2003), safety behavior (Garavan & OBrien, 2001), and expert ratings (Zohar, 1980) but contradict those examining accident history (Lee & Harrison, 2000; Vredenburgh, 2002). This contradiction is important to safety climate research as many other validation attempts have focused on accident history (e.g., Mearns et al., 2003; Niskanen, 1994; O'Toole, 2002; Silva et al., 2004; Varonen & Mattila, 2000; Zohar, 2000) based on the assumption that there are hypothesized paths from climate to behavior to accidents (Zohar, 2002). This study further indicates a lack of correspondence between recorded levels of observed percent safe and recorded accident rates. In this same factory, a non-significant correlation ($r=0.95$, n.s.) was obtained (Cooper et al., 1994) when the association between the site's observed percent safe and lost time accident rate was examined. Instead, large associations for absenteeism rate ($r=0.75$, $p<.01$) and machine "downtime" ($r=0.67$, $p<.01$) were demonstrated. Such findings indicate that there is a weak link between *actual* behavior and officially recorded accident rates. Rather, other more generalized organizational factors that affect safety appear more closely linked to accident rates (Gershon et al., 1999; Mearns et al., 2003; Van Vuuren, 2000; Varonen & Mattila, 2000) and it may be these that the majority of safety climate studies are tapping into rather than levels of safety behavior *per se*. It could be argued, therefore, that where the path from safety climate to accidents is hypothesized to be via behavior (Zohar, 2002) the use of accident rates for validation in safety climate research may not be appropriate. In these instances, it would appear that the relationship between actual behavior and accident rate needs to be established prior to establishing any relationships between safety climate scales and accident rate. However, when both organizational and safety climate are examined together (Silva et al., 2004), accident rate might be more appropriate as the effects of the measured organizational climate can be partialled out. A further note of caution when using accident rate for validation, however, is to recognize that as well as their being a potential for lack of reliability (Zohar, 1980), historical fact (i.e., the past reality) is inevitably compared with present perceptions. In other words, "apples" and "oranges" are being compared.

4.3.1. Magnitude of climate and behavior change

In general, the null hypothesis that changes in safety behavior will not be reflected in equal changes in safety climate scores was supported. Only one department's change scores matched reasonably well. Positive changes in safety climate scores within three departments were not matched with equally positive changes in behavior. However, two departments exhibiting less change in safety climate scores were accompanied by greater changes in behavior. This latter finding supports other evidence indicating that behavioral improvement programs of the type described here may lead to behavior change without any noticeable change in attitudes (McKnight & McPherson, 1986). An alternative explanation is that these results are confounded by changes in the behavioral safety checklists, in that different sets of behaviors were being monitored at the different time points. However, it could be argued that the safety climate of the facility also changed over time as remedial actions were completed and safety behavior improved. Moreover, safety climate measures are designed to capture such changes. The one constant in organizational life is change. As such, safety climate and safety behavior is dynamic and always in a state of flux, which supports the regular use of safety climate surveys. Nonetheless, whatever the explanation there has been a long held assumption that safety climate scores provide an accurate representation of actual safety performance. These study results provide a timely warning that safety climate scores do not necessarily reflect actual levels of safety behavior or safety performance *per se*, suggesting that the use of multiple performance indicators might be warranted to validate safety climate measures. This would also offer the advantage of providing further insights into the underlying relationships between employee perceptions about safety and safety performance *per se*.

4.4. Study limitations

Although this study adds to the extant safety climate literature by obtaining apparent safety climate–behavior relationships, the study has two specific limitations. First, when attempts have been made to establish climate–behavior relationships this study has relied heavily on correlation coefficients with a limited number of cases resulting from the aggregation of department scores. This has obviously diminished statistical power, which in turn inhibits any causal interpretations of the results. Following in the footsteps of Mearns et al. (2003), a perusal of the pattern of correlations as well as the absolute values suggest the obtained degrees of association between the two variables are truly indicative of the magnitude of relationship. However, correlations do not indicate cause and effect, merely that there is a statistical relationship. Thus, we do not know whether the exhibited levels of safe behavior influenced employee's perceptions of safety climate, or whether the prevailing safety climate influenced actual safety behavior. However, the scale concerning perceptions of the importance of safety training does appear to be a predictor of actual safety behavior. Further research will establish whether this result is generalizable to other sample populations.

Second, the study was conducted in one manufacturing facility, in one industry, in one country. Thus the results may not generalize to other industries, manufacturing facilities, or countries, although Zohar and Luria's (2003) oil refinery results suggest that changes in observed behavior do lead to changes in perceptions of safety climate. They did not, however, attempt to correlate the behavior scores with the safety climate scores to establish any relationships.

This study is the first in a series of four conducted during the 1990s by the first author where a safety climate measure was distributed twice (at least 12 months apart), while a behavioral safety initiative was also implemented. These will be reported in the near future, but research by others is also required to replicate these results in different industries utilizing different safety climate instruments and outcome measures.

5. Impact on industry

The finding that safety climate perceptions will not necessarily match actual levels of safety performance strongly suggests that industry should focus its *primary* safety improvement effort on changing unsafe situations and conditions as well as people's safety behavior at all organizational levels, rather than concentrating on improving people's attitudes, beliefs, and perceptions about safety. It is reductions in the frequency of unsafe behaviors and their antecedents (i.e., unsafe conditions or situations) that reduce the opportunity for accidents to occur, not perceptions about how safety is operationalized. Support for this viewpoint comes from empirical evidence that shows that hypothesized paths from attitudes and beliefs (i.e., climate perceptions) to behavior to accidents and injuries are weak (Lund & Aarø, 2004). This is not to downplay the importance of perceptions about safety climate for improving safety performance. In accordance with Carroll (1998) the role of such perceptions is very important in highlighting where system and physical changes are required within an organization, as well as safety related behaviors at all hierarchical levels. As such, all organizations should regularly survey their prevailing safety climate to identify potential issues.

6. Summary

This study has established an empirical link between a limited set of safety climate perceptions and actual safety behavior. It has also demonstrated how complex the overall relationship is: changes in climate perceptions do not necessarily reflect changes in levels of behavioral safety performance. Equally, changes in safety behavior are not necessarily reflected in safety climate perceptions. Such results challenge many of the assumptions that have typified previous research. This degree of complexity requires that much more research is undertaken on the relationship between actual safety performance and safety climate, making use of a variety of safety performance outcome variables rather than relying primarily on self-report instruments that lead to "*cul de sacs*." After nearly 25 years research we are still under "starters orders" with a long way to go before the field can truly begin to progress in a meaningful way to the benefit of its major stakeholders: employees and industry.

References

Arboleda, A., Morrow, P. C., Crum, M. R., & Shelley II, M. C. (2003). Management practices as antecedents of safety culture within the trucking industry: similarities and differences by hierarchical level. *Journal of Safety Research*, *34*, 189–197.

Barling, J., Loughlin, C., & Kelloway, E. K. (2002). Development and Test of a Model Linking Safety-Specific Transformational Leadership and Occupational Safety. *Journal of Applied Psychology*, *87*, 488–496.

Barrick, M., & Mount, M. (1991). The big 5 personality dimensions and job performance: A meta analysis. *Personnel Psychology*, *44*, 1–25.

Bausell, R. B. (1986). *Practical Guide to Conducting Empirical Research*. London: Harper Collins.

Brown, K. A., Willis, P. G., & Prussia, G. E. (2000). Predicting safe employee behavior in the steel industry: Development and test of a sociotechnical model. *Journal of Operations Management*, *18*, 445–465.

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*, 445–470.

Budworth, N. (1997). The development and evaluation of a safety climate measure as a diagnostic tool in safety management. *IOSH Journal*, *1*, 19–29.

Byrom, N., & Corbridge, J. (1997, 22–24 Sept.). A tool to assess aspects of an organisations health & safety climate. *Proceedings of International Conference on Safety Culture in the Energy Industries.* University of Aberdeen.

Cabrera, D. D., & Isla, R. (1998). The role of safety climate in a safety management system. In A. Hale, & M. Baram (Eds.), *Safety Management: The challenge of change*. Oxford: Elsevier.

Carroll, J. S. (1998). Safety culture as an ongoing process: Culture surveys as opportunities for enquiry and change. *Work & Stress*, *12*, 272–284.

Cheyne, A. J. T., Cox, S., Oliver, A., & Tomas, J. M. (1998). Modelling safety climate in the prediction of levels of safety activity. *Work & Stress*, *12*, 255–271.

Cohen, J. (1977). *Statistical Power Analyses for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Earlbaum.

Cohen, J. (1988). *Statistical Power Analyses for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum.

Cohen, J. M. (2002). Measuring safety performance in construction. *Occupational Hazards*, *64*(6), 41–44.

Cook, J. D., Hepworth, S. J., Wall, T. D., & Warr, P. B. (1993). *The Experience of Work: A Compendium and Review of 249 Measures and Their Use*. London: Academic Press.

Cooper, M. D. (2000). Towards a Model of Safety Culture. *Safety Science*, *32*(6), 111–136.

Cooper, M. D., & Phillips, R. A. (1994). Validation of a Safety Climate Measure. *The British Psychological Society, Annual Occupational Psychology Conference*. Birmingham Metropole, Jan 3–5.

Cooper, M. D., Phillips, R. A., Sutherland, V. J., & Makin, P. J. (1994). Reducing Accidents with Goal-setting & Feedback: A field study. *Journal of Occupational and Organisational Psychology*, *67*, 219–240.

Cox, S. J., & Cheyne, A. J. T. (2000). Assessing safety culture in offshore environments. *Safety Science*, *34*, 1–3.

Cox, S. J., & Cox, T. R. (1991). The structure of employee attitudes to safety: A European example. *Work & Stress*, *5*, 93–106.

Cox, S. J., & Flin, R. (1998). Safety culture: philosophers stone or man of straw. *Work & Stress*, *12*, 189–201.

Coyle, I. R., Sleeman, S. D., & Adams, N. (1995). Safety Climate. *Journal of Safety Research*, *26*, 247–254.

Cronbach, L. J. (1951). Coefficent alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Dedobbeleer, N., & Beland, F. (1991). A safety climate measure for construction sites. *Journal of Safety Research*, *22*, 97–103.

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.

Donald, I., & Canter, D. (1994). Employee attitudes and safety in the chemical industry. *Journal of Loss Prevention in the Process Industries*, *7*, 203–208.

Drexler, J. (1977). Organizational Climate: Its homogeneity with organizations. *Journal of Applied Psychology*, *62*, 38–42.

Duff, A. R., Robertson, I. T., Phillips, R. A., & Cooper, M. D. (1994). Improving safety by the modification of behavior. *Construction Management and Economics*, *12*, 67–78.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: identifying the common features. *Safety Science*, *34*, 177–193.

Fuller, C. W., & Vassie, L. H. (2001). Benchmarking the safety climate of employees and contractors working within a partnership arrangement: A case study in the offshore oil industry. *Benchmarking*, *8*, 413–430.

Garavan, T. N., & OBrien, F. (2001). An investigation into the relationship between safety climate and safety behaviors in Irish organizations. *Irish Journal of Management*, *22*, 141–170.

Gershon, R. R. M., Karkashian, C. D., Vlahov, D., Kummer, L., Kasting, C., Green-McKenzie, J., et al. (1999). Compliance with universal precautions in correctional health care facilities. *Journal of Occupational and Environmental Medicine*, *41*(3), 181–189.

Glendon, A. I., & Litherland, D. K. (2000). Safety climate factors, group differences and safety behavior in road construction. *Safety Science*, *39*, 157–188.

Glendon, A. I., & Stanton, N. (2000). Perspectives on safety culture. *Safety Science*, *34*, 193–214.

Grindle, A. C., Dickinson, A. M., & Boettecher, W. (2000). Behavioral Safety and Research in Manufacturing Settings: A Review of the Literature. *Journal of Organizational Behavior Management*, *20*, 29–68.

Guldenmund, F. (2000). The nature of safety culture: a review of theory and research. *Safety Science*, *34*, 215–257.

Guzzo, R. A., Jette, R. D., & Katzell, R. A. (1985). The effects of psychologically based interventions on worker productivity. *Personnel Psychology*, *38*, 275–291.

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*, 145–161.

Hofmann, D. A., & Stetzer, A. (1996). A cross-level investigation of factors influencing unsafe behaviors and accidents. *Personnel Psychology*, *41*, 307–339.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of Meta-Analysis*. London: Sage.

International Atomic Energy Agency [IAEA] (1988). *Basic Safety Principles for Nuclear Power Plants*. Safety Series No 75-INSAG-3. Vienna: International Nuclear Safety Advisory Group.

Joyce, W., & Slocum Jr., J. W. (1984). Collective climates: Agreement as a basis for defining aggregate climates in organizations. *Academy of Management Journal*, *27*, 721–742.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31–36.

Lee, T., & Harrison, K. (2000). Assessing safety culture in nuclear power stations. *Safety Science*, *34*, 61–97.

Lund, J., & Aarø, J. E. (2004). Accident prevention. Presentation of a model placing emphasis on human, structural and cultural factors. *Safety Science*, *42*(4), 271–324.

McAfee, R. B., & Winn, A. R. (1989). The use of incentives/feedback to enhance workplace safety: A Critique of the literature. *Journal of Safety Research*, *20*, 7–19.

McDonald, N., & Ryan, F. (1992). Constraints on the development of safety culture: A preliminary analysis. *Irish Journal of Psychology*, *13*, 273–281.

McKnight, A. J., & McPherson, K. (1986). Evaluation of peer intervention training for high school alcohol safety education. *Accident Analysis and Prevention*, *18*, 339–347.

Mearns, K., Flin, R., Gordon, R., & Fleming, M. (1998). Measuring safety climate in offshore installations. *Work & Stress*, *12*, 238–254.

Mearns, K., Whitaker, S. M., & Flin, R. (2003). Safety climate, safety management practice and safety performance in offshore environments. *Safety Science*, *41*, 641–680.

Miller, K. I., & Monge, P. R. (1986). Participation, satisfaction, and productivity: A meta- analytic review. *Academy of Management Journal*, *29*, 727–753.

Muchinsky, P. M. (2004). *Psychology applied to work* (7th ed.). Florence, KY: Thomson/Wadsworth.

Mullen, B. (1989). *Advanced BASIC Meta-Analysis*. Hillsdale, NJ: L. Erlbaum.

Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behavior. *Safety Science*, *34*, 99–109.

Newman, J. (1975). Understanding the organizational structure-job attitude relationship through perceptions of the work environment. *Organizational Behavior and Human Performance*, *14*, 371–397.

Niskanen, T. (1994). Assessing the safety environment in work organization of road maintenance jobs. *Accident Analysis & Prevention*, *26*, 27–39.

Norusis, M. J. (1985). *SPSSx: Advanced Statistic Guide*. New York: McGraw-Hill.

O'Toole, M. (2002). The relationship between employees' perceptions of safety and organizational culture. *Journal of Safety Research*, *33*, 231–243.

Prussia, G. E., Brown, K. A., & Willis, G. P. (2003). Mental models of safety: do managers and employees see eye to eye? *Journal of Safety Research*, *34*, 143–156.

Reber, R. A., Wallin, J. A., & Duhon, D. L. (1989). Safety programs that work. *Personnel Administrator*, *34*, 66–69.

Rousseau, D. M. (1985). Issues of level in organizational research: multi-level and cross-level perspectives. *Research in Organizational Behavior*, *7*, 1–37.

Rummel, R. J. (1970). *Applied Factor Analysis*. Evanston, IL: Northwestern University Press.

Rundmo, T. (1992). Risk perception and safety on offshore petroleum platforms—part II: Perceived risk, job stress and accidents. *Safety Science*, *15*, 53–68.

Rundmo, T. (1994). Associations between safety and contingency measures and occupational accidents on offshore petroleum platforms. *Scandinavian Journal of Work Environment and Health*, *20*, 128–131.

Schneider, B., & Bowen, D. (1985). Employee and customer perceptions of service in banks: Replication and extension. *Journal of Applied Psychology*, *70*, 423–433.

Silva, S., Lima, M. L., & Baptista, C. (2004). OSCI: an organisational and safety climate inventory. *Safety Science*, *42*, 205–220.

Strickoff, R. S. (2000). Safety performance measurement: Identifying prospective indicators with high validity. *Professional Safety*, 45.

Sulzer-Azaroff, B., Harris, T. C., & Blake-McCann, K. (1994). Beyond training: Organizational management techniques. *Occupational Medicine: State of the Art Reviews*, *9*, 321–339.

Tarrants, W. E. (1980). *The Measurement of Safety Performance*. New York: Garland STPM.

Thompson, R. C., Hilton, T. F., & Witt, L. A. (1998). Where the safety rubber meets the shopfloor: A confirmatory model of management influence on workplace safety. *Journal of Safety Research*, *29*, 15–24.

Van Vuuren, W. (2000). Cultural influences on risks and risk management: Six case studies. *Safety Science*, *34*, 31–45.

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 and Prevention*, *21*, 761–769.

Von Eye, A., & Schuster, C. (1998). *Regression Analysis for Social Sciences*. New York: Academic Press.

Vredenburgh, A. G. (2002). Organizational safety: Which management practices are most effective in reducing employee injury rates? *Journal of Safety Research*, *33*, 259–276.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta- analysis. *Journal of Applied Psychology*, *75*, 315–321.

Williamson, A. M., Feyer, A. M., Cairns, D., & Biancotti, D. (1997). The development of a measure of safety climate: The role of safety perceptions and attitudes. *Safety Science*, *25*, 15–27.

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

Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on microaccidents in manufacturing jobs. *Journal of Applied Psychology*, *85*, 587–596.

Zohar, D. (2002). Safety Climate: Conceptual and measurement Issues. In James C. Quick, Lois E. Tetrick, & Lennart Levi (Eds.), *Handbook of Occupational Health Psychology*. American Psychological Association (APA).

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*, 567–577.

A Registered Safety Practicioner (RSP), A Fellow of the Institute of Occupational Safety & Health (IOSH) and a Chartered Psychologist, **Dominic Cooper** is an associate professor of safety education and a visiting professor in industrial / organizational psychology at Indiana University, Bloomington. Dr. Cooper holds a BSc from the University of East London, an MSc from the University of Hull and a PhD from the University of Manchester Institute of Science & Technology. His research focuses on safety culture, safety management systems, behavioral safety and safety climate. He has published in Safety Science, Professional Safety, Journal of Occupational and Organizational Psychology, Leadership and Organisational Development Journal, European Review of Applied Psychology and others.

Robin Phillips is a Behavioral Safety consultant with Marsh UK Ltd. A Member of the Institute of Occupational Safety & Health (IOSH), he holds a Masters in Construction Management from the University of Manchester Institute of Science & Technology. His research interests lie in behavioral safety, Safety climate and Human Error. He has published in the Journal of Occupational and Organizational Psychology, Leadership and Organisational Development Journal, European Review of Applied Psychology and others.