Safety Science 111 (2019) 49–66

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Criterion-related validity of the cultural web when assessing safety culture

M.D. Cooper[a,*], M. Collins[b], R. Bernard[b], S. Schwann[c], R.J. Knox[d]

[a] BSMS Inc., 6648 East State Road 44, Franklin, IN 46131, USA
[b] 2725 Rocky Mountain Avenue, Suite 120, Loveland, CO 80538, USA
[c] 13131 Lake Fraser Drive, Calgary, AB T2J 7E8, Canada
[d] 6648 East State Road 44, Franklin, IN 46131, USA

ARTICLE INFO

Keywords:
Safety culture assessment
Cultural web
Organisational safety
Criterion-related validity
Safety culture models

ABSTRACT

The purpose of the safety culture construct is to reduce organisational and occupational accidents. However, researchers have struggled to develop validated 'measures' of safety culture, that unequivocally link cultural traits with actual safety performance.

Johnson's (1992) [Johnson, G., (1992). Managing strategic change—strategy, culture and action. *Long Range Planning*, *25*(1), pp.2] qualitative cultural web tool was adapted to simultaneously produce quantitative effectiveness ratings of an organisation's current safety arrangements for impacting personnel's safety-related behaviour. Data was collected at 15 safety culture workshops across North America over three-weeks. The population sample comprised 700 personnel, divided into 110 respondent groups. Data were examined from two perspectives: *Within* the cultural web topics (Routines, Stories, Symbols, Influences, Values, Structures & Measures); and specific safety culture topics (Profit before safety, Culture of Fear, Safety Leadership, Compliance, Competency, Communication, Lessons Learned) derived from thematic content analysis *across* the cultural web topics.

The overall safety culture was shared and stable. Cronbach's Alpha (0.845) indicated reliability. Criterion-related validity between the organisation's Total Recordable Incident Rates (TRIR) for the cultural web topics ($r = 0.488$, $p < 0.01$) and specific safety culture topics ($r = 0.417$, $p < 0.01$) was found. Multiple regressions against specific incident records returned adjusted R^2 criterion-related validity coefficients between 0.06 and 0.45. Both perspectives confirmed the criterion-related validity of the cultural web tool, albeit stronger relationships tended to be obtained from the safety culture topics. The study results reinforce the conclusion that the tool is a reliable and valid method that can help companies reduce organisational and occupational incidents and improve their safety culture.

1. Introduction

The safety culture term is a construct used to explain how internal organisational social environments directly influence organisational risk practices that could lead to personal injuries or catastrophic process safety disasters (Antonsen, 2017). Its purpose is to improve organisational and occupational safety, by preventing low frequency, high severity events such as Chernobyl, Bhopal, Piper Alpha, Texas City, Deepwater Horizon, etc., as well as high frequency, lower impact events (i.e. personal injuries, etc.). First introduced in 1986 by the International Nuclear Safety Advisory Group (INSAG, 1986) and defined in 1991 by the International Atomic Energy Agency (IAEA) as '*that assembly of characteristics and attitudes in organisations and individuals which establishes that, as an overriding priority, [nuclear power] safety issues receive attention warranted by their significance*', the safety culture construct has become extremely important to regulators e.g. HSE, 2005; CANSO, 2008; EUROCONTROL/FAA, 2008; USDOT, 2011; PSAN, 2011; NRC, 2012; OSHA, 2013; BSEE, 2013). as well as organisations concerned with improving their safety performance to reduce incidents.

1.1. Theoretical models of safety culture

How an organisation approaches the task of improving its safety culture depends in large part on the theoretical safety culture model(s) adopted. Favoured by social scientists, the *interpretative approach* (e.g. Schein, 1983, 1990; Johnson, 1992) states the organisation is the culture, where 'cultural' realities are socially constructed solely by the organisation's membership. The emphasis of the interpretative

* Corresponding author.
E-mail addresses: info@bsms-inc.com (M.D. Cooper), Mike.Collins@nutrien.com (M. Collins), Ron.Bernard@nutrien.com (R. Bernard), Sarah.Schwann@nutrien.com (S. Schwann), info@bsms-inc.com (R.J. Knox).

https://doi.org/10.1016/j.ssci.2018.09.013
Received 8 March 2018; Received in revised form 4 June 2018; Accepted 23 September 2018
0925-7535/

approach is on gaining an in-depth understanding of the prevailing cultural influences (i.e. assumptions & attitudes) affecting people's behaviour. Conversely, the *functionalist approach* is favoured by managers and practitioners (the owners of safety culture) who view culture as a variable to be engineered to suit the prevailing circumstances to affect performance by addressing management system faults, people's safety related behaviour, risk-assessments and decision-making (Cooper, 2018).

During the period 1986–2000 three influential models of safety culture were developed to guide theory, research and practice. Guldenmund's (2000) *interpretative* three-layered organisational culture framework views 'culture' as a pattern of basic assumptions, invented, discovered, or developed by a group as it learns to cope with its problems of external adaptation and internal integration. This pattern of assumptions is considered to be valid and is taught to new members as the correct way to perceive, think, and feel in relation to those problems. From this perspective, organisational culture is not the overt behaviour or visible artefacts one might observe in a company; rather it is the assumptions that underlie the values and determine not only behaviour patterns, but architecture, office layout, dress codes, etc., (Schein, 1983). In accordance with these views, Guldenmund proposes that the safety culture construct has three layers: the bottom layer is comprised of core basic assumptions which are unconscious and unspecified (invisible) where suppositions about safety are not articulated but are taken for granted as the basis for argument or action. The middle layer, predicated on the core basic assumptions, reflects espoused beliefs and values which are operationalised as relatively explicit and conscious attitudes whose target is hardware (safety controls), software (effectiveness of safety arrangements), people (functional groups), and people's safety-related behaviours. Artefacts on the top layer are the manifestation of the previous two layers, which reflect all those visible safety objects (e.g. PPE, inspection reports, safety posters, etc.), from which it is asserted it would be difficult to comprehend an entity's safety culture (Schein, 1992). The basic assumptions are thought to differ for executives, engineers, and operators, which means the overall organisational safety culture is comprised of different sub-cultures. The emphasis of this approach is on understanding these basic assumptions and their meaning to the organisation's membership and changing these to improve performance. There is some indirect anecdotal evidence to support the model in the safety arena (e.g. Nielsen, 2014) and some statistical evidence in marketing (Homburg & Pflesser, 2000).

Cooper's (2000) *functionalist* reciprocal model, based on Bandura's (1977) Social Learning Theory, highlights that safety culture is a *product* (Schein, 1992) of multiple goal-directed interactions between internal psychological factors, overt behaviour(s), and situational workplace aspects. In this model, the prevailing organisational safety culture is reflected in the dynamic reciprocal relationships between: members' perceptions about, and attitudes towards, the operationalisation of organisational safety goals; members' day-to-day goal-directed safety behaviour; and the presence and quality of the organisation's safety systems and sub-systems to support the goal-directed behaviour. Formally adopted by the American Petroleum Institute (2015) and the American National Standards Institute (ANSI), the reciprocal safety culture model is supported by large-scale studies on accident prevention (e.g. Lund & Aarø, 2004) and safety culture (e.g. Fernández-Muñiz et al., 2009; Cooper, 2008; Lefranc et al., 2012).

Reason's (1998) *functionalist* approach equates safety culture with an 'informed culture', where members of the organisation understand and respect the hazards facing their operations and are alert to the many ways in which the system's defences can be breached or bypassed. In short, an informed culture is one in which people, at all levels do not forget to be afraid; they know where the 'edge' is without having to fall over it. To be informed requires that there is a reporting culture, which in turn relies on the presence of a centralised safety information system that collates and analyses data from incidents, near-miss reports, and other sources (behavioural observations, workplace inspections, etc.), and translates that information into knowledge, so that it can be widely disseminated (e.g. Carthey et al., 2001). This requires a learning culture where there is willingness and competence to draw the right conclusions from the safety information system. Based on this, a flexible culture is required where there is the will to implement major reforms when the need is indicated. However, the reporting culture is itself dependent on a just culture (how an organisation handles blame and punishment for actual or perceived transgressions). Reason asserts that trust lies at the heart of any safety culture. Reason's model is also supported by evidence (e.g. Collinson, 1999; Saji, 2003; Pluye & Hong, 2014).

Regardless of philosophical approach (i.e. interpretative or functionalist), each of the safety culture models have attempted to provide an actionable framework, and each has been influential in the sense that researchers, regulators and industry have made use of them in some empirical and/or practical capacity.

1.2. Assessing safety culture

Although scholars recommend using a triangulation of assessment methods such as audits (Grote and Künzler, 2000), qualitative focus group exercises (Buchan, 1999), and behavioural observations (Cox & Cheyne, 2000), the most common method for assessing organisational safety culture is via cross-sectional perceptual surveys (e.g. Clarke, 2006; Goodheart & Smith, 2014; Leitão & Greiner, 2015). Through a series of pre-determined questions targeting various safety-related topics, surveys typically measure staff perceptions about how safety is being managed at a particular moment in time (Byrom & Corbridge, 1997).

In occupational safety there are an almost infinite number of characteristics that can influence safety performance, and hence the prevailing safety culture. Previous work from both academe (Flin et al., 2000) and examinations of the results of public enquiries (Cooper & Finley, 2013) into process safety disasters (e.g. Deepwater Horizon, Texas City), identified six main topics reflecting important contributors to a safety culture. These were: [1] management/supervision, [2] safety systems, [3] risk, [4] work pressure, [5] competence, and [6] procedures and rules. Typically, these characteristics are contained in modern safety management systems (e.g. OSHA (S) 18001:2007; ANSI-Z10: 2012; ISO45001: 2018) implemented in many countries. Although these topics are found in many perceptual safety surveys, Cooper's (2016) safety culture review found such surveys typically exhibit non-existent to weak relationships to *actual* safety outcomes (e.g. safety behaviour, adverse safety incident records). After almost three decades of research, it seems sensible, therefore, to seek valid alternative methods for assessing safety culture.

A precursor to Guldenmund's (2000) model of safety culture, Johnson (1992) developed a qualitative practical 'cultural web' tool based on an amalgamation of both Schein's (1990) and Hofstede's (1990) culture models to assess an organisation's culture. 'In three layers, this examines: first, any unshared underlying unstated assumptions – this is 'the what' (bottom layer); next, espoused beliefs and values reflected in justifications for behaviour – this is 'the why' (middle layer); and lastly, behaviours and artefacts – this is 'the how' (top layer). These are reflected in visible organisational behavioural patterns. Johnson divided the latter into rituals and routine practices, stories told, symbols used, power relationships, organisation structures, and controls. As such, Johnson's cultural web topics are linked to the organisation's political, symbolic and structural aspects that reveal the mechanisms for change. Buchan (1999) used the cultural web extensively, with different groups in many countries, to assess safety culture in the offshore petrochemical industry. Seemingly well received by company personnel, no criterion-related validation against actual safety performance, such as lost-time incidents or other safety indicators, was reported. Biggs et al., (2010) argue that adverse incident metrics provide the most accurate measure of the state of a safety culture, but this does not 'locate' safety culture *per se*. Therefore, any valid safety culture assessment method should show a strong statistical relationship to incident rates in any given organisation, at any moment in time. The aim of this action research study (Bradbury-Huang, 2010) was to test the criterion-related validity of the cultural web tool when used as a safety culture assessment method against records of adverse safety incidents reported in a North American conglomerate.

2. Method

2.1. Sample

A major North American conglomerate had experienced rapid organic growth, accompanied by acquisitions and mergers, to become a leader in its field. However, the histories of the various business units acquired meant challenges were being experienced in safety performance which the organisation wished to address. Over a three-week period in August-September 2016, the organisation invited participants from its various business units to attend one of 15 one-day safety culture assessment workshops, mostly conducted off-site in hotels, being held across North America. In total, 700 direct employees attended, in batches of between 30 and 50 people. With a total worldwide employee population of 15,500, using a 95% confidence interval, the sample size is sufficient to generalise to all the organisations 12,000 North American employees, albeit with an error margin of plus or minus 3.5% in any scores. Attendees at each workshop were divided, primarily by job function, into groups of five. Groups tend to be more accurate in anonymous controlled feedback situations (Dalkey, et al.,1969), problem-solving (Laughlin et al., 2006), and decision-making (Michaelsen et al., 1989), all of which was pertinent to the entire cultural web exercise conducted. The full spectrum of the organisation's job functions was involved, inclusive of corporate executive leadership, plant leadership, regional managers, safety, health & environment (SHE), human resources, information technology, operational excellence, corporate support services, agricultural and transport operations, mining, maintenance, rail terminals and warehousing. In total, 110 delegate groups participated.

2.2. Assessment measure

Over the course of a full day, each group of attendees completed three cultural web assessment workbooks: [1] Where are we now? [2] Where do we want to be? and [3] How are we going to get there? This manuscript is limited to analysis and discussion of the 'Where are we now?' section only.

Delegate groups were required to identify and grade the effectiveness of the safety activities, processes, and systems they experience in their daily work, and any underlying assumptions they held about these, using a 10-point rating scale for each of the cultural web topics outlined below. This section took approximately two to three hours to complete, as the groups discussed each of the topics, wrote their responses, and jointly agreed an effectiveness score for each activity, process, system or assumption they identified. Johnson's (1992) cultural web topics, organised by the psychological, behavioural and situational elements of Cooper's (2000) safety culture model, that each delegate group discussed and rated were:

2.2.1. Psychological aspects of safety culture

1. 1 Stories about Safety: What stories do people tell each other when talking about safety and how effectively do they influence what people do in safety?

Stories are a mechanism by which the organisation's past is outlined and particular ways of behaving are legitimised. As such, stories are devices for informing people what the company and workers consider to be important in the organisation (e.g. Keyton, 2014).

2. **Underlying assumptions –** *A summary description of the underlying reasons for doing the identified safety activities.* People's underlying assumptions are thought to reveal how they truly perceive safety. Although these may be totally different from expectations, they are important as they are thought to determine how people approach safety in daily life (e.g. Johnson, 1992).

Respondents were asked to identify and rate three *specific* assumptions about each of the seven cultural web topics (e.g. assumptions about routines, assumptions about stories, etc.), and three separate *global* assumptions (i.e. about safety *per se*). The three effectiveness scores for each topic were averaged to provide an overall effectiveness score. In total, this meant there were eight discrete sets of assumptions available for analysis (seven specific assumptions and one global overall assumption).

2.2.2. Behavioural aspects of safety culture

3. Safety Routines: What routine activities are currently undertaken in your work place that focus on safety?

The routine ways that people behave make up 'the way we do things around here' which reflects the common practical definition of organisational [safety] culture (e.g. Deal & Kennedy, 1982).

4. Safety Values Displayed: How does your facility show it genuinely values good safety?

Many of the best performing companies adhere to a set of deeply ingrained core safety values that guides all of its actions in safety matters; they serve as one of its fundamental cultural cornerstones, with people's safety- related behaviour being a reflection of the company's safety values, beliefs and attitudes (e.g. Reiman, 2007).

2.2.3. Situational aspects of safety culture

5. Safety Symbols: What safety symbols are used to publicise SHE, and how effectively do they work?'

Symbols such as logos, titles, clean uniforms, hard hats, and the language used, or in-house sayings, become a short hand representation of the nature of the organisation. These symbols usually convey messages beyond their functional purpose. For example, T-shirts awarded for so many accident free days conveys the organisation's pleasure at the fact that no accidents occurred, rather than providing new clothing for employees. (e.g. Johnson, 1990; Lesch et al., 2009).

6. Influences on Safety: Who or what influences safety in your workplace?

This element explores the nature by which people, organisational factors (e.g. goals, systems) and the physical working environment exert an influence on people's everyday behaviour. Typically, managerial decisions, co-worker's behaviour and the safety of the working environment are likely to be associated with the organisation's core underlying assumptions and beliefs (e.g. production comes before safety) (e.g. Cooper, 2015).

7. Safety Structures: What safety management system structures do you use in your workplace?

The presence and quality of formal organisational safety structures depicts and emphasises what the company believes is important to control and mitigate risks, but unless they are used they have no effect (e.g. McDonald et al., 2000; Armstrong & Laschinger, 2006; Zeng et al., 2007).

8. Monitor, Measure & Reinforce Safety: How does your facility measure, monitor and reinforce safety in your workplace?

Given that what gets measured gets done, what a company feels is important to measure, monitor and reinforce, is the behaviour it will see (e.g. Petersen, 1996). In the best performing companies, a basket of lagging and leading indicators deliberately linked to its safety vision and achievement strategy are used, to ensure that all organisational members are continually focused on safety (e.g. Cooper, 2002; Yates et al., 2005; Reiman & Pietikäinen, 2012).

2.2.4. Effectiveness rating scale

For each of the above topics, delegate groups were asked to *qualitatively* identify any particular SHE process, system or activity that they experienced in regard to safety in their daily work lives (e.g. routinely attending toolbox talks), and any assumptions they held. Both the British HSE (2001) and Biggs et al., (2010) argue safety culture measures should focus on effectiveness to ascertain the influence a safety culture characteristic has on safety performance. As such, each delegate group was asked to *quantitatively* rate the effectiveness of the activities, processes, systems and assumptions they identified using a 1-10 effectiveness grading scale. Ten-point scales provide more options to help avoid restriction of range in response data (Dawes, 2008). It was emphasised to all delegate groups that the effectiveness rating was primarily concerned with the impact of the activity, process, system or assumption on theirs and their colleague's daily safety-related behaviour (i.e. did it have an influence on people's daily safety behaviour, and if so, to what degree). The effectiveness scale used was:

1. **Low** – Not effective at all in influencing people's safety behaviour
3. **Fairly low** – Moderate effects can be seen
5. **Average** – Some effects are obvious
7. **Good** – Usually influences what we do
9. **High** – Strongly influences everything we do
10. **Very High** = Very effective in influencing people's safety behaviour

In summary, the assessment exercise used a mixed-methods approach comprised of *qualitatively* identifying any safety-related activities, processes or systems they experienced at work within a particular cultural web topic, and *quantitatively* grading each of these by using the effectiveness rating scale.

2.3. Outcome measures

Occupational safety outcome indicators are the means by which organisations measure the results, effects or consequences of activities carried out in the context of a programme related to accident prevention, preparedness and response. They are designed to measure whether actions taken are achieving the intended results (OECD, 2003).

These indicators do not necessarily represent reality but are an attempt to reflect the truth in the form of multiple and different forms of data (Mearns, 2009). Safety effectiveness is traditionally monitored via lagging 'after the event' measurements such as accident and injury/incidents rates (Lingard et al., 2013). This is partly due to: regulatory requirements (e.g. OSHA); monitoring the effectiveness of risk controls; providing lessons learned so as to avoid any repeat events in the future; facilitating the trending of salient issues over a period of time (usually years); and, providing real-time monitoring of the prevailing safety culture (Biggs et al., 2010).

Accidents are thought to be random events that have a statistically low probability of occurrence. If monitored, they provide useful information on the functioning and failure of safety barriers, rather than reflecting an organisations' safety level *per se* (Lofquist, 2010, Reiman & Pietikäinen, 2012). This is because a low accident rate does not guarantee that safety risks are being controlled or that work-related injuries or diseases will not occur in the future. Although accident rates may be a valid (or true) indicator of current or past safety effectiveness, their reliability as a predictor of future events is dubious (Mengolinim & Debarberis, 2008). Accidents by their nature are due to a particular set of sometimes complex circumstances coming together at a particular point in time (Mearns, 2009). Moreover, it is very difficult, if not impossible, to predict the actual severity of outcome of an accident, even when the complex circumstances are similar. Safety management, therefore, needs a continuous focus on lagging indicators of past and current outcomes, including deficiencies and incidents, and a basket of 'leading' indicators targeting salient technical, organizational and human factors to drive safety forward (Reiman & Pietikäinen, 2012).

The study organisation had gone beyond minimum legal requirements with their various lagging indicators, routinely monitoring different categories of incidents reflecting different severity levels to provide different perspectives on system or barrier failures (i.e. risk-controls). Various computer-based reporting systems were available *in-situ* to facilitate incident reporting, with follow-up investigations being conducted when deemed applicable. In descending order of severity, the lagging safety performance outcome indicators used in this study were:

Incident Categories Routinely Monitored	***Definition***
Serious Injuries & Fatalities (SIFs)	– Life-threatening or life-altering events that could lead to the death of the affected individual or result in a permanent or long-term impairment or loss of use of an internal organ, body function, or body part.
Number of Lost-time incidents (LTI)	– Injuries in which the injured worker cannot return to work on the following workday
Restricted Work Cases (RWC)	– Injuries in which the injured person cannot return to the same task that was performed prior to the accident.
Number of Recordable Injuries (REC)	– An incident that results in death, days away from work, restricted work or transfer to another job, medical treatment beyond first aid, or loss of consciousness
Number of First-Aids	– Injuries that are typically minor injuries that can be treated without the need of a medical physician. In most cases these types of injuries are treated on the jobsite and the worker is expected to return to work with a minimum delay
Potential Serious Injuries & Fatalities (PSIFs)	– An event that potentially could have resulted in a life-threatening or life-altering event
Number of Near-Misses	– An unwanted, unplanned event that did not result in an actual injury, but the potential for an injury existed.
Composite Indices	

Total Recordable Incident Rate (TRIR)	–	The number of employees per 100 full-time employees that have been involved in a recordable injury or illness. These include fatalities, lost work day cases, restricted work cases, and medical treatment cases
Lost-Time Incident Rate (LTIR)	–	The number of lost time incidents per 100 full-time employees in any given time frame

In general terms, the advantages of lagging indicators include being relatively easy to collect and understand, and when based on standard formulae, facilitate the computation of standardised rates per 100 employee work-hours per year, that provide composite measures to enable benchmarking or comparative analyses (De Cieri et al., 2012). The disadvantages of using accidents as an outcome indicator of safety effectiveness include under-reporting, variability in recording, relative infrequency, and the involvement of chance factors unrelated to the inherent safety of an organisation (Clarke & Robertson, 2008). Despite such potential confounds, in the absence of other objective safety measures, this study used these safety outcome indicators as independent criterion variables with which to test the concurrent validity of the cultural web assessment tool.

2.4. Data processing

The completed documentation for each workshop was couriered overnight from each workshop location to the first author's offices for transcription, collation and analysis of data. Transcription for part one of the assessment exercise took approximately one week per workshop, depending on the number of delegate groups and responses. Independent quality checks were conducted to ensure accuracy of transcription. For example, the total number of comments within each cultural web topic in the original paper versions for a workshop were counted and compared to the number contained in the electronic versions, by a different team member than the original transcriber. Checks were also made by randomly comparing the wording of the electronic transcription against the wording in the original written version. Specific organisational terms and their meanings were also clarified by the participating organisation. In some instances, where handwriting was difficult to read, the transcribers would collectively reach a consensus, and flag it as such. These were subsequently re-checked by the first author to ensure the authenticity of the written content and its effectiveness score. In total, across the 110 respondent groups there were approximately 8800 written responses identifying various activities, processes, systems or assumptions. Each of these had a corresponding effectiveness rating score to be used for computation and statistical analysis.

2.4.1. Content analysis within a thematic framework

To provide insights into actual safety issues, *within* each cultural web topic, using a thematic framework (Ritchie et al., 2013), the responses were analysed for content and allocated to one of seven preordained safety culture topics previously identified from public enquiries into process safety disasters (e.g. Cooper & Finley, 2013; Cooper, 2016). The idea was to provide in-depth focus to guide the organisation's safety culture journey to help avoid any catastrophic events. The allocation of items, and their accompanying effectiveness score to the seven safety culture topics, was checked and re-checked by independent members of the data-processing team to ensure consistency. A few problematic items were discussed by the whole data-processing team until consensus was reached. The seven common safety culture topics, aligned with the work of Flin et al. (2000), contained in Table 1 were:

Table 1
Common safety culture topics x Psychological, Behavioural and Situational Aspects.

	Topic (Flin et al., 2000)	Topic (Cooper & Finley, 2013)	Identified Activities, Processes & Systems
Psychological Aspects			
1.	Work Pressure	Profit before Safety	Production versus Safety conflict
2.	Management/ Supervision	Culture of Fear	Just Culture (rewards & discipline)
Behavioural Aspects			
3.	Management/ Supervision	Safety leadership	Leadership activity
4.	Procedures/Rules	Compliance	Rules and Procedures
Situational Aspects			
5.	Safety Systems	Communications	Safety Communication processes
6.	Competence	Competency	Training and Competence processes
7.	Safety Systems	Lessons Learned	Incident Reporting, Investigation & Analysis

Theming these safety culture topics, after the event, provided the advantage of minimising any potential bias in responses, as participants were responding to the main cultural web topics *per se*. The essential difference between the two sets of effectiveness rating scores, is that the cultural web scores reflected *within* topic effectiveness ratings (e.g. routines), whereas the specific safety culture topics reflected *between* cultural web topics: i.e. aggregate effectiveness ratings calculated *across* all seven cultural web topics (excluding assumptions). In combination all the topics assessed in this work (cultural web and safety culture) covered key features contained within all three safety culture models. For example, Guldenmund's (2000) 'assumptions, behaviours and artefacts', Cooper's (2000) psychological, behavioural and situational aspects, and Reason's (1998) reporting, learning, informing & just cultures'. As such, as a secondary exercise, the results of the cultural web exercise offer a rare opportunity to simultaneously gain insights into the theoretical validity of each model.

2.5. Data analysis

All the delegates' effectiveness rating scores were entered into statistical software (SPSS – Statistical Package for the Social Sciences) for computation and analysis. Initially, this was used to assess the measure's internal reliability, and to produce a 'mean average' effectiveness rating for each topic. Subsequently tests of the method's external validity were conducted against the organisation's year-to-date safety-related incident records (eight months), available at the time of the assessment, making use of multiple regression to try and identify which features were associated with the various type of safety incidents.

3. Results

3.1. Reliability

Reliability refers to the consistency of measurement (e.g. the effectiveness ratings). If a measurement method is not consistent, it is impossible to conclude that any scores accurately measure the domain of interest (e.g. safety culture).

Reliability analysis is often viewed as a first-step in a measurement method's validation process. If a measure is not reliable, in turn, it cannot be valid. Therefore, any observed correlation coefficient between a criterion (e.g. safety outcome) and independent variable (e.g. Cultural Web topic) will tend to underestimate the true magnitude of validity (Worthen et al., 1999). Thus, the meaningfulness of any validity coefficients obtained in this study is dependent upon the "reliability" of both the criterion (e.g. incident rates) and the cultural web assessment tool.

Table 2
Criterion related reliability coefficients * = significant at the < 0.05 level; ** = significant at the < 0.01 level.

LTI	−0.782**								
RWC	−0.418**	0.943**							
REC	−0.418**	0.771**	0.895**						
FA	0.726**	−0.086	−0.024	0.367**					
PSIFs	0.797**	−0.300*	−0.322*	0.600**	0.754**				
NM	−0.066	−0.135	−0.289*	0.259*	0.591**	0.614**			
TRIR	−0.418**	0.687**	0.649**	0.665**	−0.096	0.228	0.131		
LTIF	−0.782**	0.389**	0.849**	0.136	−0.759**	−0.986**	−0.779**	0.242	
	SIFs	LTI	RWC	REC	FA	PSIFs	NM	TRIR	LTIF
Average Reliability Estimate (Rho)	0.60*	0.52	0.63*	0.53	0.38	0.62*	0.39	0.37	0.68*

3.1.1. Criterion reliability

Due to the non-parametric nature of incident data, Spearman's Rho (Spearman, 1910), an ordinal alternative to Cronbach's alpha, was used to assess the reliability of the study's criterion variables (i.e. incident categories). A Rho > 0.60 is considered adequate reliability, while Rho's between > 0.25–0.60 are considered moderate. The obtained Spearman's Rho reliability coefficients are shown in Table 2. Unlike *linear* correlations (e.g. Pearson) which show that as one variable increases or decreases in relation to another, Spearman's Rho show the strength of a *monotonic* relationship between paired data. i.e. a relationship where one variable never increases or decreases as a result of another variable increasing or decreasing. The closer Rho is to ± 1 the stronger the monotonic relationship. Similarly, the more a variable measures different features, the greater the possibility of negative average covariance among items and hence negative reliability coefficients. For example, the positive monotonic reliability coefficients for the SIFs – PSIFs (Rho = 0.797, p < .01) and SIFs-First-Aids (Rho = 0.726, p < .01) show these highly related incident categories share common features. Conversely, as shown by the negative reliability coefficients, all the other incident types measure something else. As such, on this basis it could be argued SIFs are an entirely different class of incident (which is exactly what they are!).

To provide an aggregated estimate of the overall criterion reliability of each incident category, the Spearman Rho correlations were subjected to a Fisher z transformation. The resulting z-values were averaged and converted back to an r value. This process is known to lead to reduced magnitude coefficients (Silver & Dunlap, 1987) but produces less bias than directly averaging the correlation values (Corey et al., 1998). As such, the estimated aggregate reliability coefficients are considered conservative. The transformed correlations show acceptable criterion-related reliability (Rho ≥ 0.60) for Serious Injuries & fatalities (SIFs), Restricted Work Cases (RWC), potential SIFs (PSIFs) and Lost-Time Incident Frequency (LTIF). Moderate aggregate reliability coefficients were obtained for the number of Lost-Time Incidents (LTI), Recordables (REC), First-Aids (FA), Near-Misses (NM) and the Total Recordable Incident Rate (TRIR). Thus, any validity coefficients obtained in this study are likely to be much lower than they would have been, had the incident categories had higher internal consistency.

3.1.2. Cultural web reliability

A reliability coefficient of 0.70 is considered the minimum acceptable for a psychometric measurement method, while somewhere between 0.80 and 0.90 is considered desirable (e.g. Peterson, 1994; Tavakol & Dennick, 2011). A Cronbach's Alpha test was conducted on the eight cultural web topics. The obtained coefficient (r = 0.845) indicates there was a high degree of consistency among the group's responses, suggesting the cultural web assessment results are consistent and reliable. A Cronbach's Alpha test was also run on the derived safety culture topics. The obtained coefficient (r = 0.848) was similar to that obtained from the cultural web topics, albeit marginally higher, again confirming the consistency of the effectiveness ratings across the participant workgroups. As such, the mixed-methods assessment used in this study has shown a desirable level of reliability for use in the real-world.

3.2. Criterion validity

No objective work-related criterion variable can ever undergo a completely satisfactory empirical test of its adequacy, as it is subject to many biasing contextual factors such as type of work, size of work-groups, working environment, competence of personnel, leadership style, etc. Consequently, a criterion variable must be logically justifiable as valid in its own right (Brogden & Taylor, 1950). The various types of incident records used in this study are those typically reported in the vast majority of companies around the globe by company personnel via their company incident reporting systems, and as such are the best available estimate of safety-related system and barrier failures, and their severity of impact (i.e. safety effectiveness). Logically, if a report is entered into the incident reporting system an adverse event occurred, therefore, the use of incident records is a valid criterion variable that can be justified. A such, it is inferred they have construct validity. Nonetheless, given the moderate reliability coefficients for some incident types obtained in this study, questions could be raised concerning the degree to which they truly reflect the organisations incident rates as they could be contaminated by under-reporting (e.g. Clarke & Robertson, 2008). If levels of under-reporting are significant, it could contaminate any subsequent validation analysis using the Cultural Web assessment results (Brogden & Taylor, 1950).

To accurately estimate levels of under-reporting requires knowledge of the true number of incidents experienced by employees relative to the number actually reported (Probst & Estrada, 2010). Many studies using safety incident data have used self-report surveys to provide such estimates, even though they are known to be contaminated by common method variance (Podsakoff et al, 2003) social desirability responding (Paulhaus, 1989), and participants problems recollecting past events (Liao et al., 2001). Others have compared workers compensation costs to OSHA records to provide estimates of organizational under-reporting to the authorities, even though workers' compensation records may also be contaminated by underreporting (Parker et al., 1994). Thus, any retrospective estimates of the extent of under-reporting using the aforementioned methods are likely to be inaccurate. This poses the question 'how can we know what we don't know?'

In the absence of asking questions about the delegates accident experience during the workshops, attempts to assess the extent of any under-reporting used chi-square analysis to test the "goodness of fit" between the number of year-to-date reported incidents and the expected number of incidents. Chi-square is the sum of the squared difference between the observed and expected data, divided by the expected data in all possible categories. The chi-square test was used because the random nature of incidents means they are distribution-free (i.e. The distribution of incidents won't follow a normal bell curve) and because it assesses the null hypothesis, which states that there is no significant difference between the expected and observed result. The resulting p value is the *probability* that the deviation of the observed from that expected is due to chance alone (i.e. no other forces are exerting an effect). Should a difference emerge, it could either indicate the

Table 3
Chi-square analysis results.

	SIFs	LTI	RWC	REC	FA	PSIFs	NM	TRIR	LTIF
Chi-Square	13.71	1.09	0.66	8.31	10.05	7.03	2.95	2.73	1.09
df	1	6	6	7	9	7	10	9	6
Asymp. Sig.	0.000	0.982	0.995	0.306	0.347	0.426	0.983	0.974	0.982

presence of under-reporting, or simply reflect the difference in risk represented by the activities of those concerned in the various business units, workshop locations, employment categories and job functions (Gawande et al., 2003).

The chi-square results shown in Table 3 revealed a significant difference ($p < 0.000$) between observed and expected SIF values. No significant differences were found for any other types of incident category, suggesting any variation in the data for these criterion variables is entirely due to chance, and that under-reporting was not a systemic problem.

A further examination of the SIF data revealed that the significant difference for the SIF reports were associated with the high-risk nature of maintenance and operations compared to the lower-risk nature of corporate services and other administrative functions: in other words, the nature of the work. Thus, no solid evidence could be found indicating there was systemic or significant levels of under-reporting. As such, it could be concluded, with a high degree of confidence, that the incident data in this study represents a reasonably true reflection of incident occurrence.

Just to be sure, the ratios of minor to serious injury reports (Rebbitt, 2014) were also computed and examined along the lines suggested by Heinrichs triangle, which postulates that the severity distribution of occupational accidents is relatively constant due to common causes (e.g. Marshall et al., 2018). In general, the magnitude of the ratios should be lower at the top, and higher at the bottom of the triangle. Using the year-to-date reported incident records (8 months), Table 4 reveals this severity distribution held true with the exception of First-Aids (FA) and Potential Serious Injuries & Fatalities (PSIFS). The lower magnitude ratios for First-Aids were contrary to expectations, suggesting there was some level of under-reporting, although the extent is unknown. Conversely, the Chi-Square analysis above suggests it could simply be due to chance. Nonetheless, any validity relationships of the cultural web scores to First-Aids should be treated with some degree of caution.

On the other hand, PSIFs are a special class of reported near-miss events that potentially could to lead to an immediate life-altering or life-threatening outcome. Not usually represented in the Heinrich triangle, this ratio provides useful information regarding the likelihood of an actual SIF occurring, which is the purpose of recording PSIFs. A relatively new type of safety metric (Wachter & Ferguson, 2013) it is unknown whether or not these are under-reported, although the Chi-Square analysis would suggest otherwise.

In sum, although there does appear to be a possibility that some under-reporting occurred, it is likely this was restricted mainly to relatively minor First-Aid events, which may be caused by issues such as a perceived low severity of injury and the potential negative reactions of others (Tucker et al., 2014). Overall, taking both analytic attempts as a whole, with the exception of first-aids, it would appear that the reported incident categories are not unduly affected by under-reporting, and therefore, provide a reasonably accurate reflection of actual incident occurrence.

Table 4
Incident Category Ratios in accordance with Heinrich's Triangle.

Ratios	SIF	LTI	RWC	REC	FA	PSIF
SIF						
LTI	13:1					
RWC	15:1	1:1				
REC	42:1	3:1	3:			
FA	34:1	3:1	2:1	1:1		
PSIF	16:1	1:1	1:	0.4:1	0.5:1	
Near-Miss	786:1	61:1	54:1	19:1	23:1	49:1

Table 5
Overall means & standard deviations for cultural web topic effectiveness ratings.

Cultural Web Topic	n	$\bar{x}$	s.d.
Psychological Aspects			
Stories about Safety	108	3.79	2.37
Summary Global Assumptions about Safety	81	5.34	3.34
Specific Assumptions – Routines	110	5.82	2.67
Specific Assumptions – Stories	110	5.09	3.37
Specific Assumptions – Symbols	110	5.77	2.67
Specific Assumptions – Influences on Safety	110	5.48	3.08
Specific Assumptions – Values Displayed	110	5.57	3.27
Specific Assumptions – Safety Structures	110	4.82	2.97
Specific Assumptions – Monitor, Measure & Reinforce	110	4.57	3.10
Behavioural Aspects			
Safety Routines	110	5.04	1.47
Safety Values Displayed	108	5.06	2.23
Situational Aspects			
Safety Symbols	109	4.03	1.76
Influences on Safety	110	4.42	2.18
Safety Structures	110	4.75	1.72
Monitor, Measure and Reinforce Safety	108	5.00	1.87
Overall Cultural Web Effectiveness Score	*110*	*4.68*	*1.44*

3.3. Mean average scores – cultural web topics

Shown in Table 5, the mean 'Effectiveness' rating was 4.68 (s.d. = 1.44) out of a possible 10. The three highest scoring cultural web topics were all related to the underlying assumptions. The highest was assumptions about safety routines ($\bar{x}$ = 5.82; s.d. = 2.67), followed by assumptions about safety symbols ($\bar{x}$ = 5.77; s.d. = 2.67), and assumptions about the safety values displayed ($\bar{x}$ = 5.57; s.d. = 3.27). The mean score for the global summary assumptions was ($\bar{x}$ = 5.34; s.d. = 3.34). Safety values displayed (i.e. behaviour) was rated at 5.06 (s.d. = 2.23), with safety routines (i.e. behaviour) rated at 5.04 (s.d. = 1.47). The effectiveness of the organisation's monitoring, measuring and reinforcement of safety received a rating of 5.00 (s.d. = 1.87), safety structures was rated as 4.75 (s.d. = 1.72), influences on safety was rated as 4.42 (s.d. = 2.18) and safety symbols as 4.03 (s.d = 1.76). The lowest scoring topic was stories about safety (i.e. psychological aspect) with an effectiveness rating of 3.79 (s.d = 2.37).

3.4. Mean average scores – safety culture topics

Shown in Table 6, the overall mean average effectiveness rating for the safety culture topics was 4.59 (s.d. = 1.38). Compliance to rules and procedures (i.e. behaviour) with a rating of 5.48 (s.d = 1.50) was the highest scoring. Profit before safety (i.e. values, beliefs & attitudes) received an aggregate rating of 4.72 (s.d. = 2.02), followed by communications (i.e. situational) with a rating of 4.70 (s.d. = 1.69). Safety leadership (i.e. behaviour) received an effective rating of 4.63 (s.d. = 2.02), with lessons learned (i.e. situational) rated as 4.46

Table 6
Overall means & standard deviations for safety culture topics effectiveness ratings.

Safety Culture Topics	n	$\bar{x}$	d.
Psychological Aspects			
Profit before Safety	109	4.72	2.02
Culture of Fear	109	4.25	2.12
Behavioural Aspects			
Safety Leadership	109	4.63	2.02
Compliance to Rules/Procedures	110	5.48	1.50
Situational Aspects			
Communications	109	4.70	1.69
Competency	102	3.85	1.95
Lessons Learned	105	4.46	1.97
Overall Safety Culture Effectiveness Score	*110*	*4.59*	*1.38*

(s.d. = 1.97). A culture of fear (i.e. values, beliefs & attitudes) received a rating of 4.25 (s.d. = 2.12), with competency (i.e. situational) receiving the lowest rating at 3.85 (s.d. = 1.95).

3.5. Statistical difference tests

The demographic information collected during each assessment exercise included workshop location, respondent group's business unit (e.g. corporate, operations), employment category (e.g. manager, supervisor or employee), and job function (e.g. human resources, etc). One-way analysis of variance (ANOVA's) was used to test for any statistical differences between each group effectiveness ratings, with the Scheffe test used as a post-hoc procedure to ensure any differences obtained were real. Presented in Table 7, no statistically significant differences were found in effectiveness scores between any of the groups for any of the cultural web topic or the safety culture topics. Although there were clearly some differences, these non-significant results indicate, with a high degree of confidence, that to all intents and purposes the effectiveness ratings of the organisation's safety culture were *shared* by all and were *stable* – both of which are key elements of organisational culture (Hofstede et al., 1990, p. 311).

3.6. Criterion-related validity

For an assessment method to be useful, in conjunction with being reliable, it must also be valid: i.e. it must be related to some other external measure of performance in the same domain (e.g. safety performance). Criterion-related validity is indicated when measures on the predictor (e.g. cultural web) and criterion variables (e.g. safety incidents) are correlated, and the strength of the correlation coefficient (r) substantially supports the extent to which the assessment method estimates performance on each criterion (Waltz et al., 2005). As a general rule, the higher the validity coefficient, the more beneficial the assessment method. Many scholars recommend criterion-related validity coefficients (r) > 0.45 (DeVon et al., 2007), whereas the US Dept. of Labor considers validity coefficients above 0.35 to be very beneficial (Saad et al., 1999).

3.6.1. Correlations

To test for criterion-related validity, the delegate group's overall effectiveness ratings for the cultural web topics were initially correlated with the organisation's year-to-date Total Recordable Incident Rates (TRIR), specific only to the different regional business units involved in the assessment exercise. TRIR is an aggregate measure encompassing many different types of safety incident within one metric. The obtained

Table 7
Significance statistics for inter-group differences in effectiveness ratings.

Cultural Web Topics	Workshop Location	Business Unit	Employment Category	Job Function
Psychological Aspects				
Stories about Safety	0.321	0.510	0.468	0.871
Global Assumptions about Safety	0.998	0.331	0.129	0.235
Specific Assumptions – Safety Routines	0.803	0.303	0.169	0.871
• Specific Assumptions – Stories about safety	0.527	0.112	0.072	0.104
• Specific Assumptions – Safety Symbols	0.612	0.507	0.114	0.625
• Specific Assumptions – Influences on Safety	0.279	0.606	0.174	0.089
• Specific Assumptions – Values Displayed	0.156	0.103	0.115	0.170
• Specific Assumptions – Safety Structures	0.657	0.207	0.317	0.109
• Specific Assumptions – Monitor, Measure & Reinforce	0.436	0.607	0.278	0.675
Behavioural Aspects				
Safety Routines	0.666	0.968	0.484	0.970
Safety Values Displayed	0.204	0.063	0.114	0.720
Situational Aspects				
Safety Symbols	0.977	0.086	0.460	0.151
Influences on Safety	0.759	0.449	0.345	0.954
Safety Structures	0.676	0.952	0.235	0.991
Monitor, Measure and Reinforce Safety	0.900	0.313	0.567	0.528
Overall Cultural Web Score	*0.680*	*0.342*	*0.202*	*0.840*
Safety Culture Topics	Workshop Location	Business Unit	Employment Category	Job Function
Psychological Aspects				
Profit before Safety	0.498	0.328	0.908	0.940
Culture of Fear	0.926	0.827	0.321	0.364
Behavioural Aspects				
Safety Leadership	0.390	0.788	0.264	0.806
Compliance to Rules/Procedures	0.962	0.956	0.938	0.732
Situational Aspects				
Communications	0.946	0.849	0.919	0.997
Competency	0.391	0.740	0.739	0.089
Lessons Learned	0.936	0.568	0.994	0.989
Overall Safety Culture Topic Score	*0.527*	*0.524*	*0.923*	*0.912*

validity coefficient (r = 0.488, $p < 0.01$) clearly shows the aggregate effectiveness ratings were related to the organisation's aggregate incident rate. This was repeated for the safety culture topics, which again returned a validity coefficient (r = 0.417, $p < 0.01$). The strength of these two correlations with the organisation's year-to-date TRIR substantially supports the use of the cultural web as a valid safety culture assessment method (Waltz et al., 2005).

3.6.2. Multiple regression analysis

Given the strength of the two criterion-related correlational validity coefficients, the next step was to try and discover which cultural web topic(s) or specific safety culture topic(s) were associated with specific types of adverse incident records. Step-wise multiple regressions were run using the year-to-date incident statistics as the dependent variable, and the cultural web or safety culture topics as the independent variables. Any statistically significant relationship provides evidence of the topic(s) association with the various types of incident. Validity coefficients for two or more topics combined indicate that it is the incremental combination of those topics that explains the strength (R^2) of the relationship to that particular type of incident. The *coefficient of determination* (R^2) is the proportion of the variance in the dependent variable (e.g. safety incident records) that is explained from the independent variable(s) (e.g. cultural web topics). The *adjusted* R^2 is a modified version of R^2 that has been adjusted for the number of predictor variables in the regression model, which increases only if an additional independent variable improves the overall model more than would be expected by chance (Von Eye & Schuster, 1998). If an accompanying *F*-value is statistically significant (typically $p < .05$), the model explains a significant amount of variance in the outcome variable (e.g. incident records).

High correlations between the independent variables (i.e. topics) could indicate multicollinearity (inter-correlated predictors) that can cause severe computational problems (O'Brien, 2007): i.e. by over-inflating the standard errors, multicollinearity would make some variable coefficients statistically insignificant when they should be significant (i.e. a type 2 error). Without multicollinearity, those coefficients might be significant (Akinwande, Dikko, & Samson, 2015). 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). A VIF of *n* means that the variance of the estimated coefficients is *n* times higher because of the correlation between the independent variables. However, the extant literature is not clear on what VIF threshold value to use: some use a threshold value of 5 (e.g. Rogerson, 2001), while most use a VIF > 10 (e.g. Hair et al., 1995) to sequentially eliminate problematic variables from the regression equation in an attempt to reduce the collinearity and optimise the significance of the model (Akinwande et al., 2015).

3.6.3. Multiple regression – cultural web topics

The first series of step-wise multiple regressions focused solely on relationships between the cultural web topics (i.e. routines, stories, symbols, values displayed, influences, safety structures, monitor & measure, global assumptions about safety, and topic specific assumptions) and the different types of incident. Given the importance of 'underlying assumptions' in organisational culture models (e.g. Johnson, 1992; Schein, 1992; Guldenmund, 2000) Cultural Web topic-specific 'underlying assumptions' (e.g. assumptions about the safety routines) as well as the global summary of the organisation's underlying assumptions about safety, were each entered into the model as separate discreet variables. Initially, only the cultural web topics were entered. The global summary assumptions were added next. Finally, the topic-specific underlying assumptions were added. With this three-phase approach, the aim was to identify the extent to which the respondent groups effectiveness ratings of their 'underlying assumptions', influence the organisation's actual safety performance. Contained in Table 8, the criterion-related validity coefficients (*adjusted* R^2) for the Cultural Web topics ranged from 0.14 to 0.35 ($p < .05$).

The predictors for both minor and serious incidents were: values displayed; measuring, monitoring and reinforcing safety; telling stories about safety; and specific assumptions about safety stories. Multicollinearity statistics were between 1.0 and 1.51, well below the VIF thresholds of 5 or 10. With eight cultural web topics, and seven specific sets of underlying assumptions, out of a potential 15 topics, only four exhibited criterion- related validity with actual safety performance. Moreover, none were associated with the Lost-Time Incident Rate (LTIR), even though this incident category has the highest criterion reliability estimate (Rho = 0.68).

3.6.4. Multiple regression – safety culture topics

Multiple-regression analyses were conducted to examine the influence of the specific safety culture topics (independent variables) on the organisation's incident records (dependent variable). Shown in Table 9, every safety culture topic was significantly related, to some degree, to all of the organisation's incident types. The criterion-related validity coefficients (*adjusted* R^2) ranged from 0.08 to 0.45 ($p < .01$). Multicollinearity statistics were between 1.0 and 1.31, again well below the VIF thresholds of 5 or 10. The magnitude of these validity coefficients was, in general, much greater than the cultural web topics, with adjusted R^2 coefficients of 0.25 or higher being obtained for seven types of incident. Given the stronger statistical relationships, the seven specific safety culture topics appear to be more valid predictors of an organisation's safety performance than the cultural web topics, further supporting their use as the prime targets for assessing and improving safety culture (Cooper, 2016).

With the exception of restricted work cases (REC) and near-misses (NM), the topic *Culture of Fear* was related to seven types of incident records, indicating that developing a 'Just' culture (Reason, 1998) is absolutely crucial to reducing adverse safety incidents. *Profit before Safety* was related to the total Recordable Incident Rate (TRIR), the number of Lost-Time incidents (LTI), recordable injuries (REC), and restricted work cases (RWC), clearly indicating the significant impact of the safety-productivity conflict on injuries (e.g. Hamidi, et al., 2012). Similarly, *Competency* was related to serious injuries & fatalities (SIFs), the lost-time incident rate (LTIR), the number of first-aids (FA) and potential serious injuries & fatalities (PSIFs), indicating that a lack of competency exerts a significant influence on both serious and minor injuries (Aronsson, 1999). *Compliance* to rules and procedures was related to serious injuries & fatalities (SIFs), lost-time incident rate (LTIR) and the number of first-aids (FA). *Communications* was related to the number of lost-time incidents (LTI) and number of restricted work cases (RWC), while *Lessons Learned* was related to the number of recordable injuries (REC), and the number of restricted work cases (RWC).

3.7. Safety culture model testing

This study found a wide range of safety culture topics were significantly linked via criterion-related validity coefficients with various types of incident records. As such, they present an opportunity to evaluate the validity of the theoretical models of safety culture that provide the framework for much current safety culture research.

3.7.1. Organisational culture model

Contrary to organisational culture theory (Schein, 1992, Guldenmund, 2000), associated underlying assumptions appear to have little influence on actual safety performance. Behaviours (i.e. safety values displayed) and artefacts (i.e. monitor, measure & reinforce, and stories about safety) at the top level (Johnson, 1992) explained most of the associations with the various incident types. Specific 'assumptions about safety stories' was the only type of 'underlying assumption' that appeared in the regression models. These relationships were for the Total Recordable Injury Rate, in conjunction with safety values displayed, (adj. $R^2 = 0.35$; $p \leq 0.01$); the number of recordable injuries

Table 8
Criterion-related validity coefficients for the cultural web topics: DV = Dependent Variable; IV = Independent Variables.

Types of Incident (DV)	Rho	Cultural Web Topics – Predictors (IV)	Adj. R^2	F	Sig.
Total Recordable Incident Rate (TRIR)	0.37	Assumptions (Stories), Safety Values Displayed	0.35	14.99	0.01
Number of Lost-time incidents (LTI)	0.52	Safety Values Displayed	0.29	16.54	0.01
Number of Recordable Injuries (REC)	0.53	Assumptions (Stories)	0.28	21.94	0.01
Potential Serious Injuries & Fatalities (PSIFs)	0.62*	Monitor, Measure & Reinforce, Stories about Safety	0.25	9.89	0.01
Number of Near-Misses (NM)	0.39	Stories about Safety	0.19	14.99	0.01
Serious Injuries & Fatalities (SIFs)	0.60*	Monitor, Measure & Reinforce	0.14	5.84	0.02
Restricted Work Cases (RWC)	0.63*	Safety Values Displayed	0.14	7.08	0.01
Number of First-Aids (FA)	0.38	Safety Values Displayed, Assumptions (Stories)	0.14	5.24	0.03
Lost-Time Incident Rate (LTIR)	0.68*	None	–	–	–

Criterion related reliability coefficients * = significant at the < 0.05 level.

Table 9
Criterion-related validity coefficients for the safety culture topics.

Types of Incident (DV)	Rho	Safety Culture Topics – Predictors (IV)	Adj. R^2	F	Sig.
Serious Injuries & Fatalities (SIFs)	0.60*	Fear, Compliance, Competency	0.45	10.12	0.01
Number of Recordable Injuries (REC)	0.53	Profit before Safety, Lessons Learned, Fear	0.42	16.48	0.01
Total Recordable Incident Rate (TRIR)	0.37	Fear, Profit before Safety	0.41	23.18	0.01
Lost-Time Incident Rate (LTIR)	0.68*	Competency, Fear, Compliance	0.36	9.88	0.01
Number of Lost-time incidents (LTI)	0.52	Fear, Profit before Safety, Communication	0.35	9.44	0.01
Number of First-Aids (FA)	0.38	Competency, Compliance, Fear	0.26	9.50	0.01
Restricted Work Cases (RWC)	0.63*	Lessons Learned, Profit before Safety, Communication	0.24	5.52	0.01
Potential Serious Injuries & Fatalities (PSIFs)	0.62*	Competency, Fear	0.14	6.19	0.01
Number of Near-Misses (NM)	0.39	Profit before Safety	0.08	7.72	0.01

Criterion related reliability coefficients * = significant at the < 0.05 level.

(adj. $R^2 = 0.28$; $p \leq 0.01$); and, again in combination with safety values displayed, the number of First-Aids (adj. $R^2 = 0.14$; $p \leq 0.03$).

3.7.2. *Reciprocal safety culture model*

Cooper's (2000) reciprocal safety culture model highlights that safety culture is a *product* (Schein, 1992) of multiple goal-directed interactions between internal psychological factors, overt behaviour(s), and situational workplace aspects. The product is defined as '*that observable degree of effort by which all organisational members direct their attentions and actions toward improving safety on a daily basis*' (Cooper, 2002). The effectiveness ratings are to some extent a measure of this product, as they are concerned with the impact of the organisation's efforts to improve safety on people's daily safety behaviour. To test the model, SPSS was used to compute composite psychological, behavioural and situational variables by summing the effectiveness ratings for the different Cultural Web topics in the respective categories shown in Table 2.

The criterion-related validity coefficients (adjusted R^2) for the composite variables ranged from 0.01 and 0.33. Multicollinearity statistics ranged between 1.0 and 2.84, well below the VIF thresholds of 5 or 10, but higher than all other multiple regression analyses. In general, the size of the coefficients for the composite variables were generally lower than those obtained for the original cultural web topics presented in Table 5, with the exception of the associations for the number of Lost-time incidents (LTI) which was slightly higher (0.33 vs 0.29).

Shown in Table 10, the composite psychological variable (e.g. Stories, global underlying assumptions, specific underlying assumptions) did not appear in any of the regression models, with the composite behavioural and situational aspects showing some statistical relationships with the organisation's incidents. Specifically, both the behavioural and situational variables were related to the number of lost-time incidents (adj. $R^2 = 0.33$, $p < .01$), the TRIR (adj. $R^2 = 0.25$, $p < .01$), and the number of near-misses reported (adj. $R^2 = 0.11$, $p < .02$). The behavioural composite was related to *all* the different types of incident, although three were non-significant (potential serious injuries & fatalities, number of first-aids, and lost-time incident rate). Overall, these results suggest that focusing on behaviour is the best way to influence a safety culture.

A similar exercise was undertaken with the seven safety culture topics. Shown in Table 11, criterion-related validity coefficients (adjusted R^2) for the composite variables ranged from 0.01 to 0.44., with a combination of all three composite variables being significantly related to the TRIR (adj. $R^2 = 0.44$, $p < .01$) the number of recordable injuries (REC) (adj. $R^2 = 0.31$, $p < .01$) and the number of near-misses reported (NM) (adj. $R^2 = 0.09$, $p < .02$). Multicollinearity statistics were between 1 and 1.69. Jointly, the behavioural and situational composite variables were related to all other incident types. This suggest that a focus on optimising both the situation and behaviour, rather than just focusing on psychological or behavioural variables, will exert a positive influence on adverse safety incidents, the same conclusion Cooper (2016) came to in his narrative safety culture review. However, this is tempered by the fact that the safety-productivity conflict and a culture of fear may provide a negative backdrop that will prevent any improvement efforts flourishing.

Again, the magnitude of the criterion-related validity coefficients for the composite variables derived from the specific safety culture topics were generally lower than those obtained for the original safety culture topics. Nonetheless, the reciprocal safety culture model is clearly related to safety performance.

3.7.3. *Reasons (1998) safety culture model*

Reason's model of safety culture comprises five sub-cultures that interact to create the safety culture. These are just, reporting, learning, flexible and informed cultures. Combinations of four related variables measured in this study showed statistically significant links to the organisation's incidents. These were a culture of Fear (Just Culture), Lessons Learned (Reporting & Learning Cultures), Communications, and Stories about safety (Informed Culture). Clearly, Reason's (1998) notion that trust is at the heart of a safety culture was strongly supported, as it was associated with seven types of safety incident, highlighting the importance of this variable to all efforts to improve organisational safety cultures. Similarly, lessons learned, communication and stories about safety were also associated with adverse incidents, but to a much lesser degree than the culture of fear. Thus, the available data

Table 10
Cultural web composite psychological, behavioural & situational variables and incident rates.

Types of Incident (DV)	Rho	Cultural Web Topics – Predictors (IV)	Adj. R^2	F	Sig.
Number of Lost-time incidents (LTI)	0.52	Behaviour, Situation	0.33	10.37	0.01
Number of Recordable Injuries (REC)	0.53	Behaviour	0.25	19.09	0.01
Total Recordable Incident Rate (TRIR)	0.37	Behaviour, Situation	0.25	9.58	0.01
Serious Injuries & Fatalities (SIFs)	0.60*	Behaviour	0.11	4.77	0.04
Number of Near-Misses (NM)	0.39	Behaviour, Situation	0.11	4.50	0.02
Restricted Work Cases (RWC)	0.63*	Behaviour	0.10	5.38	0.03
Potential Serious Injuries & Fatalities (PSIFs)	0.62*	Behaviour	0.03	1.60	n.s.
Number of First-Aids (FA)	0.38	Behaviour	0.02	2.32	n.s.
Lost-Time Incident Frequency (LTIF)	0.68*	Behaviour	0.01	1.37	n.s.

DV = Dependent Variable; IV = Independent Variables.
Criterion related reliability coefficients * = significant at the < 0.05 level.

Table 11
Safety culture composite psychological, behavioural & situational variables and Incident Rates.

Types of Incident (DV)	Rho	Safety Culture Topics – Predictors (IV)	Adj. R^2	F	Sig.
Total Recordable Incident Rate (TRIR)	0.37	Psychology, Behaviour, Situation	0.44	17.99	0.01
Number of Lost-time incidents (LTI)	0.52	Behaviour, Situation	0.33	8.79	0.01
Number of Recordable Injuries (REC)	0.53	Psychology, Behaviour, Situation	0.31	10.86	0.01
Serious Injuries & Fatalities (SIFs)	0.60*	Behaviour, Situation	0.19	4.79	0.02
Lost-Time Incident Rate (LTIR)	0.68*	Behaviour, Situation	0.12	4.11	0.02
Number of First-Aids (FA)	0.38	Behaviour, Situation	0.09	4.53	0.01
Number of Near-Misses (NM)	0.39	Psychology, Behaviour, Situation	0.09	3.45	0.02
Potential Serious Injuries & Fatalities (PSIFs)	0.62*	Behaviour, Situation	0.06	3.26	0.05
Restricted Work Cases (RWC)	0.63*	Behaviour, Situation	0.01	0.71	n.s.

DV = Dependent Variable; IV = Independent Variables.
Criterion related reliability coefficients * = significant at the < 0.05 level.

strongly support Reason's central assertion about a just culture but provide limited evidence for the other measured sub-cultures.

3.8. Summary of multiple regression analyses

Due to the large number of multiple regression results from the various analyses, a summary table was produced showing the size of the criterion-related validity coefficients for each type of adverse safety incident by topic analysis (see Table 12). This shows that the regression models for the safety culture topics returned the largest criterion-validity coefficients for seven types of *actual* adverse incident. The cultural web topics returned larger coefficients for two incident types (near-misses and potential SIFs) that *could* have led to an injury. Summing the various validity coefficients to produce an average, shows that assessing safety culture with specific safety topics accounted for about 33% more of the organisation's safety performance than the general cultural web topics.

However, the large number of multiple regressions conducted in this study may have led to type 1 errors (i.e. accepting a supposed effect or relationship that does not really exist), or type 2 errors (i.e. rejecting an effect or relationship that actually does exist). In other words, a percentage of the statistically significant and/or non-significant results might be entirely due to chance. As there is a trade-off between the two error types (i.e. controlling for one increases the chance of the other), the only way to reduce the possibility of both at the same time is to increase the sample size. Although the study sample size was 700 people, the cases used for analyses were 110 delegate groups, thus drastically reducing statistical power.

Table 12
Summary of criterion-related validity coefficients by Analysis. Shaded areas = possible type 1 or type 2 errors.

Types of Incident	Criterion Reliability Estimate – Rho	Cultural Web Topics (n = 15)	Cultural Web Topics – Composite (n = 3)	Safety Culture Topics (n = 7)	Safety Culture Topics – Composite (n = 3)
Serious Injuries & Fatalities (SIFs)	0.60*	*0.14*	*0.11*	0.45	0.19
Total Recordable Incident Rate (TRIR)	0.37	0.35	0.25	0.41	0.44
Number of Recordable Injuries (REC)	0.53	0.28	0.25	0.42	0.31
Lost-Time Incident Frequency (LTIF)	0.68*	–	n.s.	0.36	*0.12*
Number of Lost-time incidents (LTI)	0.52	0.29	0.33	0.35	0.33
Number of First-Aids (FA)	0.38	*0.14*	*n.s.*	0.26	*0.09*
Potential Serious Injuries & Fatalities (PSIFs)	0.62*	*0.25*	*n.s.*	*0.14*	*0.06*
Number of Restricted Work Cases (RWC)	0.63*	*0.14*	*0.10*	0.24	*n.s.*
Number of Near-Misses (NM)	0.39	*0.19*	*0.11*	*0.08*	*0.09*
Average Validity Coefficient		**0.22**	**0.19**	**0.30**	**0.20**

Criterion related reliability coefficients * = significant at the < 0.05 level.

To identify the results where type 1 or type 2 errors may exist, the effect sizes obtained this study were entered into BioStat's Power & Precision: Version 2 software (e.g. Cohen et al., 2001) to calculate the minimum sample size required, given the desired probability level (0.01–0.05), the number of predictors in the regression models (n = 3, 7 and 15), the actual effect sizes (0.06–0.45), and the desired level of statistical power (0.8). Those identified as problematic are italicised in Table 12: i.e. a sample size greater than the 110 study cases was needed to obtain the reported effect-sizes while excluding the possibility of a type 1 or 2 error.

This exercise showed that the cultural web topics, despite their < 0.05 significance levels, might not exhibit criterion-related validity with five criterion variables: potential and actual Serious Incidents & Fatalities (PSIFs & SIFs), Restricted Work Cases (RWC), First-Aids (FA) and Near-misses (NM). Equally, the safety culture topics may not be significantly related to PSIFs or Near-misses (NM): i.e. they may all be false positives. In terms of the cultural web topics, we can only be confident that there is an unequivocal statistically significant link to three criterion variables: The Total Recordable Incident Rate (TRIR), the Number of Recordable Injuries (REC) and the Number of Lost-time incidents (LTI). The converse was found for the Safety Culture topics: only two (PSIFs & NM) may be false positives. The remainder were all unequivocally and significantly linked statistically to the other seven criterion variables. The same results were found for the composite measures for the Cultural Web topics as above. Within the composite regressions for the safety culture topics examining psychological, behavioural and situational composites, there could be false acceptance of four significant relationships: PSIFS, FAs, NM and LTIF: i.e. a type 1 error. Equally, within both sets of composite regressions there may be false rejection (Type 2 error) of the cultural web topics with regard to PSIFS, FAs, and LTIF, and RWC for the safety culture topics.

It is recognised that the unreliability of the criterion variables will also have played its part in these findings, by reducing the magnitude of the validity coefficients without correcting for attenuation, thereby increasing the sample size required to entirely avoid type 1 and 2 errors. It is argued, however, that adverse incident data in the real world, is almost certain to be unreliable (i.e. inconsistent) due to its random nature. Despite such issues, the fact that the cultural web tool was found to be unequivocally linked statistically to the majority of reported incident categories (while acknowledging the type 1 or type 2 error concerns about others), is testament to the utility of the cultural web approach to assessing safety culture.

4. Discussion

The use of the cultural web tool (Johnson, 1992) as described, appears to provide a practical method for assessing organisational safety culture that is both reliable (r = 0.845) and externally valid (i.e. significantly related to various types of the organisation's safety incident records). Statistical evidence that unambiguously links safety culture or safety culture elements with safety performance outcomes is rare (Sorensen, 2002). This is mainly because few empirical studies have been conducted that examine the relationship between safety culture and actual safety performance metrics (Cole et al., 2013). Helping to address this shortfall, this study has described a useful method for assessing safety culture.

Different from the 'one-size fits all' approach used in perceptual safety climate surveys, which use sets of pre- determined questions (Guldenmund, 2007), the cultural web tool used a concurrent, integrative, qualitative – quantitative approach, where there is much greater insight to be gained from the combination of both qualitative and quantitative data than either form by itself (Creswell & Creswell, 2017). Groups of respondents served as informants who qualitatively described what actions their organisation was taking to improve safety, using their own terms and concepts to express their point of view (Rousseau, 1990). Respondent groups quantitatively rated the effectiveness of the impact of these self-identified activities, processes, systems and underlying assumptions on theirs and their colleague's daily safety-related behaviour. The ten-point effectiveness scale provided a common frame of reference (Wreathall, 1995) that was simple to interpret, within the cultural web topics of interest (e.g. routines, stories, symbols, values displayed, influences, structures, metrics, assumptions), and the safety culture topics (Profit before safety, a culture of fear, safety leadership, compliance, communications, competence, and lessons learned) derived from thematic content analyses across the cultural web topics.

In the absence of reliable and valid leading safety indicators (e.g. safety behaviours, visible safety leadership, etc.), real-world adverse incident records are generally the only outcome data available for validating safety culture assessments. Although incidents are considered to be random events that signal system or barrier failures, in this study attempts were made to identify levels of under-reporting and estimate the reliability (internal consistency) of the incident categories (criterion measures). Chi-square analysis of all reported incidents within their specific categories did not identify any systemic under-reporting of incidents. Conversely, calculating ratios in accordance with Heinrich's triangle did suggest there may have been some under-reporting of first-aid incidents. The aggregate reliability estimates (Rho) > 0.6 were considered adequate (Gilpin, 1993) for Serious Injuries & Fatalities (SIFs), Lost-time Incident Frequency (LTIF), Potential Serious Injuries and Fatalities (PSIFs), and Restricted Work Cases (RWC). Moderate aggregate reliability estimates (Rho ≥ 0.0.25–0.60) were obtained for the Number of Lost-time incidents (LTI), Number of Recordable Injuries (REC), Number of First-Aids (FA), Near-Misses (NM) and the Total Recordable Incident Rate (TRIR). No attempts were to made to correct the reliability coefficients for attenuation, simply because adverse incident records are considered random events and will, therefore, always involve a similar lack of reliability (Schmitt and Klimoski, 1991). Given that low reliability for either a criterion, predictor or both will decrease the magnitude of a resulting validity coefficient, those reported in this study are considered to be conservative.

Both the cultural web and safety culture topics were linked to the organisation's actual adverse incident records, albeit stronger relationships tended to be obtained from the safety culture topics. As such, the cultural web tool lends itself to assessing safety culture from different perspectives. Cultural web topics give clues about the mechanisms to change, while the safety culture topics help to identify the specific safety characteristics to change (i.e. the what and the how). However, the higher unequivocal criterion-related validity coefficients for the safety culture topics known to be implicated in industrial disasters and serious injuries (Cooper, 2016) strongly suggests these should be the ultimate targets for safety culture assessment and improvement initiatives. Nonetheless, it does appear that measuring the wider organisational culture factors measured by the cultural web topics should be the first port of call to eliminate potential respondent bias, before allocating the responses into the safety culture topics.

4.1. Theoretical issues – cultural web topics

The fact this study shows the cultural web provides reliable and valid data linked to safety performance, means it has the potential to be used by researchers to work toward the development of an overarching integrative model of safety culture that draws on Guldenmund's (2000), Cooper's (2000) and Reason's (1998) work. For example, in this study, it was found that basic underlying assumptions effectiveness ratings appear to have little influence on an organisation's adverse incident rates: Only specific underlying assumptions about the safety stories told in the organisation appeared in any of the regression models (i.e. TRIR, Recordable Injuries, First-Aids). This is contrary to the central thesis of interpretative models of organisational culture (e.g. Johnson, 1992; Guldenmund, 2000), which assert that the essence of culture lies in the set of underlying assumptions held by the organisation's membership

(Schein, 1992). Several possibilities might provide an explanation:

First, the underlying assumptions were not consistently the same for all respondent groups (Schein, 1984). However, no statistically significant differences emerged in the effectiveness ratings for any of the underlying assumptions between workshop locations, business units, employment status, or job function. This indicates the *perceived impact* of the underlying assumptions on people's daily safety behaviour were shared by all and were stable (Hofstede et al., 1990, p. 311), even if the assumptions themselves were not all identical. Moreover, the non-significant differences between all the various demographic groupings for all the cultural web topics also demonstrated consistent, shared and stable behaviour patterns with regards to safety across the organisation. It seems, therefore, that a lack of consistency can be discounted.

This begs the question "was the stability of the effectiveness ratings due to the underlying assumptions or some other factor"? Calculating the Coefficient of Determination between a cultural web topic (e.g. routines) and its corresponding underlying assumption provides some insight. Adjusted R^2 results ranged from 0.26 to 0.41 and all were statistically significant ($p \leq .01$) and did not include possible type 1 or type 2 errors. Therefore, although the basic underlying assumptions appear to overlap with the corresponding cultural web topics by 26–41%, they do not fully explain the stability of the effectiveness ratings, indicating the artefacts themselves (e.g. routines, symbols, structures, etc), or some other factor(s), are exerting a much bigger influence.

Second, the underlying assumptions are too far removed from the day-to day functioning of the organisation's processes and systems (e.g. Berson et al., 2015). Thus, the immediacy of day-to-day issues overrides or negates the influence of any underlying assumptions which are too far removed in time and space from day-to-day organisational functioning for them to be related to actual outcomes (e.g. safety performance). If this is the case, then changing people's underlying assumptions about safety (Schein, 1992; Guldenmund, 2000) will exert little influence on actual safety performance (i.e. reducing adverse incidents).

Third, the cultural web tool does not get to the 'true' basic underlying assumptions (Guldenmund, 2010, p 175), and hence no true relationship with performance would be found. The cultural web uses a focus group exercise that attempts to find out what people are doing (behaviour) and what exists (processes and systems) to control safety, what the impact of these are on people's safety behaviour (effectiveness), and the underlying reasons for doing safety the way they do (i.e. the tacit convictions/assumptions that drive action). In this study, thematic analysis was not used to derive and interpret the underlying cultural assumptions through what was written *per se* (Guest et al., 2011). Rather we took what was written as a true reflection of the respondent group's assumptions and focused solely on their effectiveness ratings of what was being done and why to influence people's safety-related behaviour(s). Perhaps using the effectiveness ratings, rather than discovering any qualitative emerging themes contained within the written responses to derive the underlying assumptions and using those as the basis for analysis, would explain the lack of relationship of the underlying assumptions to actual adverse safety incidents. This is something future research might address.

Fourth, the written underlying assumptions provided by the focus groups are accurate, and their importance to organisational outcomes has been over-stated in the extant literature (e.g. Schein, Johnson, Guldenmund). Certainly, there is a serious dearth of empirical evidence linking basic underlying assumptions to performance. That which is available (i.e. Homburg & Pflesser, 2000) showed that basic underlying values (equivalent to assumptions) and behavioural norms exerted an indirect influence on marketing behaviour, and that artefacts were critically important for guiding such behaviour. In the current study, only three cultural web artefacts and one specific set of underlying assumptions were linked to actual safety performance. Similarly, collapsing the data into composite psychological, behavioural and situational variables revealed that only the behavioural and situational variables (i.e. artefacts) were found to be linked to safety performance. It appears, therefore, that organisational safety success is not rooted in the basic underlying assumptions (the essence of culture according to Schein) *per se*. Rather, underlying assumptions may simply serve to foster the commitment necessary for a group of people to overcome safety problems (Quick, 1992) by creating appropriate behavioural norms, safety processes and systems.

4.2. Theoretical issues – safety culture topics

The importance of the safety culture topics in the 'real-world' of industrial/occupational/organisational safety cannot be underestimated. They have appeared as causal factors in many public inquiries into process safety catastrophes (e.g. Piper Alpha, Deepwater Horizon, Texas City), examinations of the causes of process safety incidents (e.g. Collins & Keely, 2003; Christou & Konstantinidou, 2012; IAEA, 2014; Wood & Gyenes, 2015) and serious injury & fatality events (e.g. Hobbs & Williamson, 2003; Haslam et al., 2005; Celik & Cebi, 2009; Wachter & Ferguson, 2013), leading Cooper (2016) to revise the reciprocal safety culture model and incorporate them. Cooper argued that focusing on these significant safety issues when undertaking safety culture assessments, should lead to much stronger relationships between the results and actual safety performance. In this study, that proved to be the case as the safety culture topics criterion-related validity coefficients were, on average, about one-third greater than those obtained for the more general cultural web topics (e.g. routines, symbols, etc), ranging between 0.08 and 0.41 (albeit the relationships with PSIFs and Near-Miss outcome measures possessed the potential for type 1 or type 2 errors).

Specific study results showed a culture of fear was related to seven adverse safety incident types, indicating that developing a 'Just Culture' is absolutely critical to reduce both serious and minor safety incidents. Supporting Reason's (1998) assertion that trust is at the heart of any safety culture, a culture of fear refers to being blamed and/or punished when things do not go as planned (e.g. Cox et al., 2006). Certainly, there is evidence that a strong culture of fear increases the number of adverse safety incidents (Singer et al, 2009), that could simultaneously lead to under-reporting of incidents (e.g. Sandberg, & Albrechtsen, 2018), with both being more likely in a compliance-based organisation, rather than one that promotes employee engagement (Khatri et al., 2009).

Similarly, the study results show the safety-productivity conflict exerts an unwanted influence on both serious and minor injuries. For a number of economic and competitive reasons productivity is often the number one priority, not safety: It is often assumed that to achieve production goals, safety has to be sacrificed. Conversely, companies also face regulatory pressures to create a safe working environment: however, it is often assumed that compliance will significantly slow down production or increase costs. Cooper & Finley (2013) suggest organisations side-step the issue by adopting the philosophy *'Safe production is understood to be, and is accepted as, the number one priority*.', an approach supported by evidence that good safety performance tends to lead to better all-round economic performance (Veltri et al., 2007; Fernández-Muñiz et al., 2009). In this study, the effectiveness ratings of safety leadership *per se* were not related to any adverse incident type. However, it could be argued that both a culture of fear and the safety-productivity conflict are visible or felt manifestations of diffused safety leadership policies and practices, and as such provide an indirect measure of the impact of safety leadership effectiveness (Kelloway & Barling, 2010) on the prevailing safety culture. Consistent with the extant literature, safety competencies were also significantly related to both serious and minor injuries (e.g. Burke et al., 2006), as were compliance (e.g. Turner et al., 2015), lessons learned (e.g. Sanders, 2015) and communications (e.g. Laurence, 2005).

Similar to the results for the cultural web topics, collapsing the data into composite psychological, behavioural and situational variables revealed that behavioural and situational variables (i.e. artefacts) were found to be statistically linked to safety performance, with the exception of restricted work cases. Unlike the cultural web topics, psychological composites of the safety culture topics were also linked to three incident types (TRIR, Recordable Injuries, and Near-Misses). Statistically significant criterion-related validity coefficients for combinations of all three composite variables ranged from 0.06 to 0.44, although the significance of those below 0.12 could be due to chance (i.e. potentially a type 1 or type 2 error). Overall, these results support Cooper's (2000) reciprocal safety culture model. They also support the notion that optimising the situation to optimise people's behaviour is the best way to improve a safety culture, rather than focusing on psychological factors such as assumptions, values and attitudes (Lund & Aarø, 2004; Cooper, 2016).

4.3. Integration of the safety culture models

On one hand, this study has shown that underlying assumptions do not appear to be important for reducing incidents (e.g. Schein, 1992; Guldenmund, 2000), with surface artefacts appearing to have greater explanatory power for reducing incidents and injuries. On the other hand, the two psychological variables, culture of fear and putting profits before safety, were shown to exert a significant influence on various incident types. In some way, these must be driven by some kind of underlying assumption, value, belief or motivation. Unfortunately, an opportunity was missed during the workshops to *quantify* any espoused values and beliefs to fully test the three-layer model, which may have produced different results. Thus, it would be premature to disregard the potential influence of underlying assumptions, espoused values, beliefs or motivations on safety performance, particularly if they provide the backdrop or context for how organisations and their membership actually approach safety.

The study results also revealed the strong influence situational and behavioural variables exert on safety performance, from whichever perspective it is examined (cultural web or safety culture topics), which accords with other accident prevention work (Lund & Aarø, 2004). Similarly, elements of Reason's (1998) model were statistically linked to actual safety performance, demonstrating its utility in the real world, albeit the 'just culture' exerted the biggest influence.

Taking all of this into consideration, the reciprocal safety culture framework appears to provide an integrative foundation for encompassing the three different safety culture models going forward, in both a practical sense and theoretically. Cooper (2016) began this integration process but excluded the taken-for-granted underlying assumptions from the psychological aspect, preferring to replace them with known values (see Fig. 1).

Nonetheless, the revised model already includes the safety culture topics shown to be linked to safety incidents in this study, which also incorporates Reason's elements, with the exception of the flexible culture piece (which in some ways is accounted for by the reciprocal nature of the model). Complete model integration would suggest simply adding the underlying assumptions to the psychological aspect of this revised reciprocal model, while retaining the values, so that future work might examine any impact they exert on the known safety culture topics and safety performance. Using the cultural web assessment tool as described would provide a mechanism that captures both qualitative and quantitative data, simultaneously facilitating both the interpretative and functionalist approaches to examining safety culture.

4.4. Practical considerations

The vast majority of purported safety culture assessment studies have actually measured safety climate (attitudes & perceptions about safety) as a proxy for measuring safety culture, with less than stellar results (Cooper, 2016). Reflecting Deal & Kennedy's (1982) notion that culture is 'the way we do things around here', the cultural web approach used in this study was anchored by respondent's ratings of the effectiveness of the organisation's existing effort to impact people's safety-related behaviour(s). Ultimately, the effectiveness of an organisation's efforts to improve safety performance (e.g. Burke et al., 2006; Robson et al., 2007) is the *key* factor in SHE. Usually determined by safety performance outcomes such as adverse safety incidents, the immediate benefit of using the cultural web as described is the reasonably accurate assessment of safety culture in the respondent's own terminology and the action planning that this allows, to help stop adverse incidents from happening in the first place. Certainly, the data obtained from the exercise described here, facilitated a four-year action plan to improve the entire organisation's safety culture.

From a safety practitioner's perspective, the cultural web assessment method is relatively simple to organise and facilitate. One immediate benefit is that the organisation's membership has the time and space to reflect upon and discuss their current safety efforts, and how these could be improved, which in turn engenders ownership of safety (e.g. Breitsprecher et al., 2014). From a research perspective the advantages are concurrent qualitative – quantitative data collection, which can be analysed separately or together, with the data lending itself to theory building in the safety culture domain. Only one-day is used per group of 50 or so respondents for facilitation of data collection. Thereafter the written materials have to be transcribed and scores collated for entry into statistical software. In general, transcription and data collation take about five days per workshop. Allowing 1–2 weeks for data analysis and report writing, the entire exercise could be completed within four weeks for a single site. In the current study, the sheer scale of the exercise meant the entire process took around 3–4 months. This is comparable to the time taken to develop (1–2 weeks), distribute (4 weeks), analyse (1 week) and report back (1–4 weeks depending on the size of the sample, locations, etc) perceptual safety climate survey results. Similarly, the cultural web exercise is more cost- effective than having safety culture researchers living within an organisation for several months as they attempt to discover what the organisation's underlying assumptions about safety might be (e.g. Guldenmund, 2010, p166), when it is so much simpler to just ask people what their underlying assumptions are about why they are doing what they are doing.

4.5. Limitations

It could be argued the cultural web tool does not truly measure safety culture, even though it was specifically developed by Johnson to examine organisational cultures (Johnson, 1992). The stratified sampling of personnel from various North American regions and business units, reflecting different levels of employment status and job functions, provided the widest possible range of views within the organisation to facilitate tests of sub-cultures (Scott et al., 2003). Thus, a strong argument for asserting this study did measure the prevailing safety culture was the statistical evidence showing the effectiveness ratings of the activities, processes and systems identified by 110 respondent groups were both shared and stable, both of which are key elements of organisational culture (Hofstede et al., 1990, p. 311). Moreover, the strength and significance of the criterion-related validity coefficients, regardless of perspective (cultural web or specific safety culture topics), further help to reinforce the conclusion that the cultural web truly does measure organisational safety culture (Biggs et al., 2010). However, it is not yet known if the same result would be found within other North American organisations, or in other regions of the world.

It may also be argued the sample size of 700 people, divided into 110 respondent groups was insufficient for the results to be generalised to the total population sample of 15,500 who work for the organisation. However, with a 95% confidence interval, power analysis suggests the sample size was sufficient, albeit accompanied by a potential error margin of plus or minus 3.5% in the effectiveness scores. Given that any

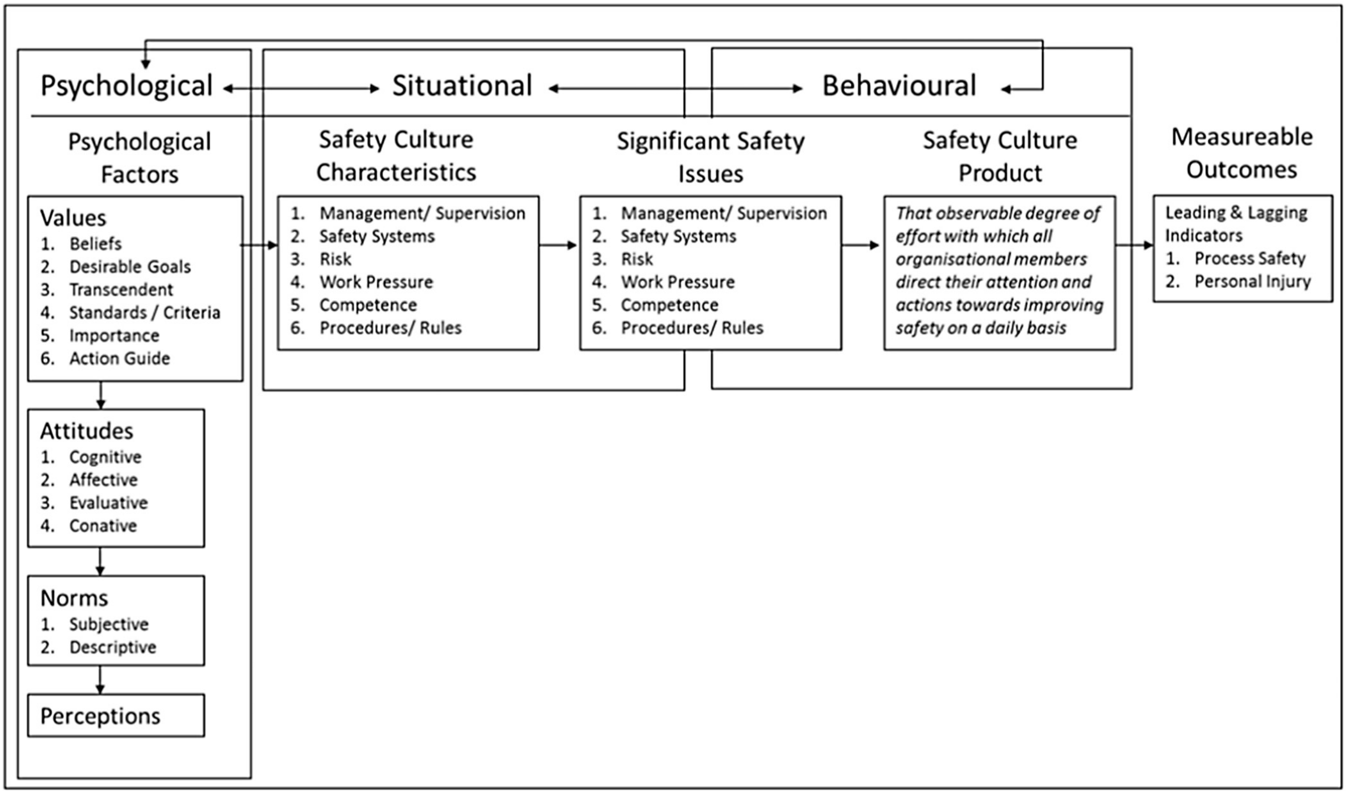

Fig. 1. Cooper's (2016) revised reciprocal model of safety culture (reprinted with permission).

adjusted R^2 coefficients obtained via multiple regression techniques correct for measurement error in both the dependent and independent variables, a high degree of confidence is held that the results were generalisable to the entire North American membership of the organisation.

Notwithstanding the above, it is clear that some of the criterion-related validities may be subject to type 1 and type 2 errors. This was due to insufficient sample sizes, combined with attenuation resulting from the moderate to adequate criterion reliability coefficients, to achieve unequivocal statistical significance. Although, type 1 and 2 errors might introduce problems in theory building in the safety culture domain, over a period of time as the study is replicated by other researchers and published, *meta*-analytic procedures can be employed to correct for reliability of the criterion and increase sample sizes (e.g. Hunter & Schmidt, 2014). Of more immediate concern, is the practical significance to industry of the actual sample sizes required to avoid type 1 and 2 errors. The problem resides in the fact that very low effect sizes require very large samples to ensure statistical significance. For example, a validity coefficient of 0.06, using three predictors (e.g. psychological, behavioural and situational composite variables), with power set at 0.8, requires a minimum sample of 262 delegate groups comprising five or so people, which translates into around 1300 delegates attending around 26 safety culture assessment workshops. Having this many delegates' may not be too much of a problem for large conglomerates or global businesses, but it presents a major problem for small to medium size enterprises who are more likely to face safety culture issues, but who many not even employ that many people. It will also significantly increase the overall costs, time and resources required to conduct the exercise. As the purpose of a safety culture assessment is to identify improvement opportunities and reduce or eliminate the likelihood of people being hurt, we would argue this purpose must take precedence over any statistical concerns regarding type 1 or type 2 errors.

There is also a possibility that the use of one instrument to measure safety culture could introduce common method variance problems (Podsakoff et al., 2003) where a general inflation across all correlations would be expected (Clarke, 2010). However, in this study, the conditions necessary to introduce common method variance (CMV) were avoided as follows: [1] CMV requires a common rater:- the effectiveness ratings in this study were achieved by consensus within each of 110 separate respondent groups, not single raters; [2] CMV can arise from the manner in which questions are presented to respondents:- topics were presented in exactly the same sequential order for each workgroup, and the only questions asked, for each broad topic (e.g. routines, stories, symbols, etc), were [a] what do you do in safety? [b] why do you do it? and [c] what is the underlying reason? without any predetermined answers whatsoever. The activities, processes, systems and underlying assumptions the respondent groups identified, and how they rated them, was entirely at their own discretion; [3] CMV can stem from the context in which questions are sequenced, and the contextual influences (time, location and media) used to measure the constructs:- the context was how respondents experienced the organisation safety improvement efforts in their daily work lives, and how effectively these influenced theirs and their colleagues safety-related behaviours; data was collected at 15 workshop locations, moderated by a single facilitator, of which there were two, mostly off-site at hotels. On some days, this meant two workshops were run concurrently in different locations. At other times, due to travel considerations, a single workshop was held by one of the two facilitators. Nonetheless, the workshop presentation material, and data recording media (i.e. worksheets) were identical for all workshops; [4] CMV often arises when the data for both the predictor (cultural web & safety culture topics) and criterion variables (safety incident records) are obtained from the same person in the same measurement context using the same item context, and similar item characteristics: – No respondent groups were asked to provide any self-report information on theirs or the organisation's incident records, as these are routinely collected and collated separately for other business and regulatory purposes by the organisation's SHE department. This study made use of those records after all the group data collection was completed, transcribed, collated and analysed in attempts to validate the results. Taking these four factors into consideration as a whole, the likelihood of CMV being introduced into the assessment process was remote in the extreme.

4.6. Conclusion

The key issue for the safety culture approach to reducing organisational and occupational accidents is to identify and demonstrate a link between an organisation's safety culture (i.e. the way cultural traits of an organisation influence safety) on the one hand, and safety performance on the other (Taylor et al., 2011). Culture is thought to influence safety in two ways: first, by providing the frames of reference through which information about risk is interpreted; the second, by influencing the way safety is formally and informally enacted within an organisation. Finding appropriate ways to 'measure' different aspects of culture has been a recurrent problem for both practitioners and researchers interested in safety culture (Antonsen, 2009). Most purported cultural assessments have used either a safety climate survey as a proxy (Cooper, 2016), or a combination of methods (interviews, behavioural observations, document analysis, audits), the results of which are not necessarily aligned with each other, nor linked to safety performance. In this study, the cultural web, using a mixed qualitative and quantitative approach that adopted the organisation memberships' point of view about the effectiveness of the current safety arrangements on impacting their behaviour, has proven to be a reliable, valid, and cost-effective method for safety culture assessment. The relatively high criterion-related validity coefficients between the cultural web and safety culture topics (traits) and the various safety incident types (performance outcomes) reported here, reinforce the conclusion that the cultural web tool is a valid predictor of safety culture.

Potential Referees

Stian Antonsen, SINTEF, Norway. Email: Stian.Antonsen@sintef.no

Ian Glendon, Griffith University, Gold Coast, Queensland, Australia. Email: i.glendon@griffith.edu.au

Kathryn Mearns, UK Health & Safety Executive, Energy Division. UK. Email: kmearns@abdn.ac.uk??

Sharon Clarke, Manchester Business School, University of Manchester, UK. Sharon.clarke@mbs.ac.uk

Acknowledgements

The authors are extremely grateful to all the back-office team members for their assistance in arranging the safety culture workshops across North America and for data transcription and collation. The authors would also thank the anonymous reviewers for their insights, which has helped to significantly strengthen the quality of the work.

Grant Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssci.2018.09.013.

References

Akinwande, M.O., Dikko, H.G., Samson, A., 2015. Variance inflation factor: as a condition for the inclusion of suppressor variable(s) in regression analysis. Open J. Statist. 5, 754–767.
American Petroleum Institute, 2015. Pipeline Safety Management Systems Standard (ANSI/API RP 1173).
Antonsen, S., 2009. Safety culture assessment: a mission impossible? J. Contingencies Crisis Manage. 17 (4), 242–254.
Antonsen, S., 2017. Safety Culture: Theory, Method and Improvement. CRC Press.
Armstrong, K.J., Laschinger, H., 2006. Structural empowerment, magnet hospital characteristics, and patient safety culture: making the link. J. Nursing Care Qual. 21 (2), 124–132.
Aronsson, G., 1999. Contingent workers and health and safety. Work, Employment Soc. 15 (3), 439–460.
Bandura, A., 1977. Social Learning Theory. Prentice-Hall, Englewood Cliffs, NJ.
Berson, Y., Halevy, N., Shamir, B., Erez, M., 2015. Leading from different psychological distances: a construal-level perspective on vision communication, goal setting, and follower motivation. Leadership Quart. 26 (2), 143–155.
Biggs, H.C., Dinsdag, D., Kirk, P.J., Cipolla, D., 2010. Safety culture research, lead indicators, and the development of safety effectiveness indicators in the construction sector. Int. J. Technol., Knowledge Soc. 6 (3), 133–140.
Bradbury-Huang, H., 2010. What is good action research? Why the resurgent interest? Action Res. 8 (1), 93–109.
Breitsprecher, K., Ndahbros, S., Hinton, J.J., Jacques, P., 2014. Organizational ownership of an interdependent HSE culture yields quantifiable HSE performance. SPE International Conference on Health, Safety, and Environment. Society of Petroleum Engineers.
Brogden, H.E., Taylor, E.K., 1950. The theory and classification of criterion bias. Educ. Psychol. Measure. 10 (2), 159–183.
BSEE, 2013. Bureau of Safety and Environmental Enforcement' Announces Final Safety Culture Policy Statement. May 9th.
Buchan, R., 1999. Safety can't wait: a global campaign for transformational change. SPE/IADC 52797. In: Proceedings of the 1999 SPE/IADC Drilling Conference, Amsterdam, March 9 ± 11.
Burke, M.J., Sarpy, S.A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R.O., Islam, G., 2006. Relative effectiveness of worker safety and health training methods. Am. J. Publ. Health 96 (2), 315–324.
Byrom, N., Corbridge, J., 1997. 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.
CANSO, 2008. Safety Culture Definition and Enhancement Process.
Carthey, J., de Leval, M.R., Reason, J.T., 2001. The human factor in cardiac surgery: errors and near misses in a high technology medical domain. Ann. Thoracic Surg. 72 (1), 300–305.
Celik, M., Cebi, S., 2009. Analytical HFACS for investigating human errors in shipping accidents. Acc. Anal. Prevent. 41, 66–75.
Christou, M., Konstantinidou, M., 2012. Safety of offshore oil and gas operations: lessons from past accident analysis. JRC Scientific & Policy Reports. European Commission.
Clarke, S., 2006. The relationship between safety climate and safety performance: a meta-analytic review. J. Occ. Health Psychol. 11 (4), 315–327.
Clarke, S., 2010. An integrative model of safety climate: linking psychological climate and work attitudes to individual safety outcomes using meta-analysis. J. Occ. Organ. Psychol. 83 (3), 553–578.
Clarke, S., Robertson, I., 2008. An examination of the role of personality in work accidents using meta-analysis. Appl. Psychol. 57 (1), 94–108.
Cohen, J., Rothstein, H., Borenstein, M., 2001. Power and Precision. Taylor & Francis.
Cole, K.S., Stevens-Adams, S.M., Wenner, C.A., 2013. A Literature Review of Safety Culture. Sandia National Laboratories, USA.
Collins, A., Keely, D., 2003. Loss of Containment Incident Analysis. HSL/2003/07.
Collinson, D.L., 1999. Surviving the rigs': safety and surveillance on North Sea oil installations. Org. Stud. 20 (4), 579–600.
Cooper, M.D., 2000. Towards a model of safety culture. Saf. Sci. 36, 111–136.
Cooper, M.D., 2002. Safety culture: understanding a difficult concept. Profess. Saf. 47 (6), 30.
Cooper, M.D., 2008. 'Risk-Weighted Safety Culture Profiling". – SPE 111823. In: The 9th International Conference on Health, Safety, Security & Environment in Oil & Gas Exploration and Production. 15–17 April 2008, Nice, France.
Cooper, M.D., 2015. Effective Safety leadership: understanding types & styles that improve safety performance. Profess. Saf. 60 (2), 49–53.
Cooper, M.D., 2016. Navigating the safety culture construct: a review of the evidence. Franklin, IN: BSMS. Retrieved from: http://behavioural-safety.com/articles/safety_culture_review.pdf.
Cooper, M.D., 2018. The safety culture construct: theory and practice. In: Gilbert (Ed.), Safety Cultures, Safety Models: Taking Stock and Moving Forward. Springer Briefs in Safety Management.
Cooper, M.D., Finley, L.J., 2013. Strategic Safety Culture Road Map. BSMS, Franklin, IN.
Corey, D.P., Dunlap, W., Burke, M., 1998. Averaging correlations: expected values and bias in combined pearson rs and Fisher's z transformations. J. General Psychol. 125, 245.
Cox, S.J., Cheyne, A.J.T., 2000. Assessing safety culture in offshore environments. Saf. Sci. 34 (1), 111–129.
Cox, S., Jones, B., Collinson, D., 2006. Trust relations in high-reliability organizations. Risk Anal. 26 (5), 1123–1138.
Creswell, J.W., Creswell, J.D., 2017. Research Design: Qualitative, Quantitative, and

Mixed Methods Approaches. Sage Publications.
Dalkey, N.C., Brown, B.B., Cochran, S., 1969. The Delphi Method: An Experimental Study of Group Opinion. Rand Corporation, Santa Monica, CA.
Dawes, J., 2008. Do data characteristics change according to the number of scale points used. Int J Market Res 50 (1), 61–77.
Deal, T., Kennedy, A., 1982. Corporate Cultures. Addison-Wesley, Reading, MA.
De Cieri, H., Shea, T., Pettit, T., Clarke, M., 2012. Measuring the leading indicators of occupational health and safety: A snapshot review. Research report#: 0612-045-R1. Dept. of Management, Monash University.
DeVon, H.A., Block, M.E., Moyle-Wright, P., Ernst, D.M., Hayden, S.J., Lazzara, D.J., Savoy, S.M., Kostas-Polston, E., 2007. A psychometric toolbox for testing validity and reliability. J. Nurs. Scholarship 39 (2), 155–164.
EUROCONTROL/FAA White Paper, 2008. Safety Culture in Air Traffic Management: A White Paper. Dec.
Fernández-Muñiz, B., Montes-Peón, J.M., Vázquez-Ordás, C.J., 2009. Relation between occupational safety management and firm performance. Saf. Sci. 47 (7), 980–991.
Flin, R., Mearns, K., O'Connor, P., Bryden, R., 2000. Measuring safety climate: identifying the common features. Safety Sci. 34 (1–3), 177–192.
Gawande, A.A., Zinner, M.J., Studdert, D.M., Brennan, T.A., 2003. Analysis of errors reported by surgeons at three teaching hospitals. Surgery 133 (6), 614–621.
Gilpin, A.R., 1993. Table for conversion of Kendall's Tau to Spearman's Rho within the context of measures of magnitude of effect for meta-analysis. Educ. Psychol. Measur. 53 (1), 87–92.
Guest, G., MacQueen, K.M., Namey, E.E., 2011. Applied Thematic Analysis. Sage.
Goodheart, B.J., Smith, M.O., 2014. Measurable outcomes of safety culture in aviation – a meta-analytic review. Int. J. Aviation, Aeronautics, Aerospace 1 (4).
Grote, G., Künzler, C., 2000. Diagnosis of safety culture in safety management audits. Saf. Sci. 34 (1), 131–150.
Guldenmund, F.W., 2000. The nature of safety culture: a review of theory and research. Saf. Sci. 34, 215–257.
Guldenmund, F.W., 2007. The use of questionnaires in safety culture research–an evaluation. Saf. Sci. 45 (6), 723–743.
Guldenmund, F.W., 2010. Understanding Safety Culture. Uitgeverij BOXPress, Nederland, pp. 167–176.
Hair Jr., J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1995. Multivariate Data Analysis, third ed. Macmillan Publishing Company, New York.
Hamidi, N., Omidvari, M., Meftahi, M., 2012. The effect of integrated management system on safety and productivity indices: Case study; Iranian cement industries. Safety Sci. 50 (5), 1180–1189.
Haslam, R.A., Hide, S.A., Gibb, A.G., Gyi, D.E., Pavitt, T., Atkinson, S., Duff, A.R., 2005. Contributing factors in construction accidents. Appl. Ergon. 36 (4), 401–415.
HSE, 2001. A guide to Measuring Health and Safety Performance, London.
HSE, 2005. A review of safety culture and safety climate literature for the development of the safety culture inspection toolkit. Research Report 367. HSE Books. London.
Hobbs, A., Williamson, A., 2003. Associations between errors and contributing factors in aircraft maintenance. Human Factors 45 (2), 186–201.
Hofstede, G., 1990. Cultures and Organisations: Software of the Mind. McGraw-Hill, London.
Hofstede, G., Neuijen, B., Ohayv, D.D., Sanders, G., 1990. Measuring organizational cultures: a qualitative and quantitative study across twenty cases. Admin. Sci. Quart. 286–316.
Homburg, C., Pflesser, C., 2000. A multiple-layer model of market-oriented organizational culture: measurement issues and performance outcomes. J. Market. Res. 37 (4), 449–462.
Hunter, J.E., Schmidt, F.L., 2014. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings, third ed. SAGE Publications Inc.
Johnson, G., 1990. Managing strategic change; the role of symbolic action. Br. J. Manage.ent 1 (4), 183–200.
Johnson, G., 1992. Managing strategic change—strategy, culture and action. Long Range Plann. 25 (1), 2.
Kelloway, E.K., Barling, J., 2010. Leadership development as an intervention in occupational health psychology. Work Stress 24 (3), 260–279.
Keyton, J., 2014. Communication, organizational culture, and organizational climate. Chapter 7. In: Schneider, B., Barbera, K.M. (Eds.), The Oxford Handbook of Organizational Climate and Culture. Oxford University Press.
Khatri, N., Brown, G.D., Hicks, L.L., 2009. From a blame culture to a just culture in health care. Health Care Manage. Rev. 34 (4), 312–322.
IAEA, 1991. Safety culture. A report by the international nuclear safety advisory group. Safety Series. 75-INSAG-4. IAEA, Vienna, Austria.
IAEA, 2014. Nuclear Safety Review 2014. GC(58)/INF/3. IEAE, Vienna.
INSAG, 1986. Summary Report on the post-accident review meeting on the chernobyl accident. Safety Series No. 75-INSAG-1. IAEA, Vienna.
Laughlin, P.R., Hatch, E.C., Silver, J.S., Boh, L., 2006. Groups perform better than the best individuals on letters-to-numbers problems: effects of group size. J. Personal. Soc. Psychol. 90 (4), 644.
Laurence, D., 2005. Safety rules and regulations on mine sites–the problem and a solution. J. Saf. Res. 36 (1), 39–50.
Lefranc, Guénolé, Guarnieri, Franck, Rallo, Jean-Marc, Garbolino, Emmanuel, Textoris, Raoul, 2012. Does the management of regulatory compliance and occupational risk have an impact on safety culture? PSAM11 ESREL 2012, Jun 2012, Helsinki, Finland, vol. 8, pp. 6514–6523, ISBN: 978-162276436-5 < hal00734322 > .
Leitão, S., Greiner, B.A., 2015. Organisational safety climate and occupational accidents and injuries: an epidemiology-based systematic review. Work Stress 1–20.
Lesch, M.F., Rau, P.L.P., Zhao, Z., Liu, C., 2009. A cross-cultural comparison of perceived hazard in response to warning components and configurations: US vs. China. Appl. Ergon. 40 (5), 953–961.
Lingard, H., Wakefield, R., Blismas, N., 2013. “If you cannot measure it, you cannot improve it”: Measuring health and safety performance in the construction industry. January 2013, 19th Triennial CIB World Building Congress, Brisbane, Aus.
Liao, H., Arvey, R.D., Butler, R.J., Nutting, S.M., 2001. Correlates of work injury frequency and duration among firefighters. J. Occup. Health Psychol. 6 (3), 229.
Lofquist, E.A., 2010. The art of measuring nothing: the paradox of measuring safety in a changing civil aviation industry using traditional safety metrics. Saf. Sci. 48 (10), 1520–1529.
Lund, J., Aarø, L.E., 2004. Accident prevention. Presentation of a model placing emphasis on human, structural and cultural factors. Saf. Sci. 42 (4), 271–324.
McDonald, N., Corrigan, S., Daly, C., Cromie, S., 2000. Safety management systems and safety culture in aircraft maintenance organisations. Saf. Sci. 34 (1), 151–176.
Marshall, P., Hirmas, A., Singer, M., 2018. Heinrich's pyramid and occupational safety: a statistical validation methodology. Saf. Sci. 101, 180–189.
Mearns, K., 2009. From reactive to proactive–Can LPIs deliver? Saf. Sci. 47 (4), 491–492.
Mengolinim, A., Debarberis, L., 2008. Effectiveness evaluation methodology for safety processes to enhance organizational safety culture in hazardous installations. J. Hazard. Mater. 155, 243–252.
Michaelsen, L.K., Watson, W.E., Black, R.H., 1989. A realistic test of individual versus group consensus decision making. J. Appl. Psychol. 74 (5), 834.
Nielsen, K.J., 2014. Improving safety culture through the health and safety organization: a case study. J. Saf. Res. 48, 7–17.
Nuclear Regulatory Commission (NRC), 2012. Safety Culture Policy statement. Dec.
O’Brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quan. 41 (5), 673–690.
OECD, 2003. Guidance on Safety Performance Indicators. Paris, France.
OSHA, 2013. Creating a Safety Culture. Retrieved from www.osha.gov/SLTC/etools/safetyhealth/mod4_factsheets_culture.html.
Parker, D.L., Carl, W.R., French, L.R., Martin, F.B., 1994. Characteristics of adolescent work injuries reported to the Minnesota Department of Labor and Industry. Am. J. Public Health 84, 606–611.
Paulhaus, D.L., 1989. Social desirability responding: some new solutions to old problems. In: Buss, D.M., Cantor, N. (Eds.), Personality Psychology: Recent Trends and Emerging Dimensions. Springer, New York, pp. 201–209.
Petersen, D., 1996. Safety by Objectives: What Gets Measured and Rewarded Gets Done. Van Nostrand Reinhold Company.
Peterson, R.A., 1994. A meta-analysis of Cronbach's coefficient alpha. J. Consumer Res. 21 (2), 381–391.
Petroleum Safety Authority Norway – PSAN, 2011. Regulations Relating To Health, Safety and the Environment in the Petroleum Activities and at Certain Onshore Facilities (The Framework Regulations).
Pluye, P., Hong, Q.N., 2014. Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. Ann. Rev. Public Health 35.
Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., Podsakoff, N.P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. 88 (5), 879–903.
Probst, T.M., Estrada, A.X., 2010. Accident under-reporting among employees: testing the moderating influence of psychological safety climate and supervisor enforcement of safety practices. Acc. Anal. Prevent. 42 (5), 1438–1444.
Quick, J.C., 1992. Crafting an organizational culture: Herb's hand at Southwest Airlines. Organ. Dyn. 21 (2), 45–56.
Reason, J., 1998. Achieving a safe culture: theory and practice. Work Stress 12 (3), 293–306.
Rebbitt, D., 2014. Pyramid power: a new view of the great safety pyramid. Professional Saf. 59 (9), 30.
Reiman, T., 2007. Assessing Organisational Culture in Complex Sociotechnical Systems: Methodological Evidence from Studies in Nuclear Power Plant Maintenance Organisations. VTT PUBLICATIONS 627. VTT Technical Research Centre of Finland.
Reiman, T., Pietikäinen, E., 2012. Leading indicators of system safety–monitoring and driving the organizational safety potential. Saf. Sci. 50 (10), 1993–2000.
Ritchie, J., Lewis, J., Nicholls, C.M., Ormston, R. (Eds.), 2013. Qualitative Research Practice: A Guide for Social Science Students and Researchers. Sage.
Robson, L.S., Clarke, J.A., Cullen, K., Bielecky, A., Severin, C., Bigelow, P.L., Irvin, E., Culyer, A., Mahood, Q., 2007. The effectiveness of occupational health and safety management system interventions: a systematic review. Saf. Sci. 45 (3), 329–353.
Rogerson, P.A., 2001. Statistical Methods for Geography. Sage, London.
Rousseau, D., 1990. Assessing organizational culture: the case for multiple methods. In: Schneider, B. (Ed.), Organizational Climate and Culture. Jossey-Bass, San Francisco, pp. 153–192.
Saad, S., Carter, G.W., Rothenberg, M., Israelson, E., 1999. Testing and Assessment: An Employer’s Guide to Good Practices. U.S. Department of Labor Employment and Training Administration, Washington, DC.
Saji, G., 2003. Safety goals in ‘risk-informed, performance-based’ regulation. Reliab. Eng. Syst. Saf. 80 (2), 163–172.
Sandberg, E., Albrechtsen, E., 2018. A study of experience feedback from reported unwanted occurrences in a construction company. Saf. Sci. 107, 46–54.
Sanders, R.E., 2015. Chemical Process Safety: Learning from Case Histories. Butterworth-Heinemann.
Schein, E.H., 1983. The role of the founder in creating organizational culture. Org. Dyn. 12 (1), 13–28.
Schein, E.H., 1984. Coming to a new awareness of organizational culture. Sloan Manage. Rev. 25 (2), 3–16.
Schein, E.H., 1990. Organizational culture. Am. Psychol. 45, 109–119.
Schein, E.H., 1992. Organisational Culture and Leadership, second ed. Jossey-Bass, San Francisco.
Schmitt, N., Klimoski, R., 1991. Research Methods in Human Resources Management.

South-Western Publishing Co, Cincinnati, Ohio.
Scott, T., Mannion, R., Davies, H., Marshall, M., 2003. The quantitative measurement of organizational culture in health care: a review of the available instruments. Health Services Res. 38 (3), 923–945.
Silver, N.C., Dunlap, W.P., 1987. Averaging correlation coefficients: should Fisher's z transformation be used? J. Appl. Psychol. 72, 146–148.
Singer, S., Lin, S., Falwell, A., Gaba, D., Baker, L., 2009. Relationship of safety climate and safety performance in hospitals. Health Services Res. 44 (2p1), 399–421.
Sorensen, J.N., 2002. Safety culture: a survey of the state-of-the-art. Reliab. Eng. Syst. Saf. 76 (2), 189–204.
Spearman, C.C., 1910. Correlation calculated from faulty data. Br. J. Psychol. 3, 271–295.
Tavakol, M., Dennick, R., 2011. Making sense of Cronbach's alpha. Int. J. Med. Educ. 2, 53.
Taylor, T.N., Eeckeleaert, L., Starren, A., Scheppingen, A.V., Fox, D., Bruck, C., 2011. Occupational Safety and Health Culture Assessment – A Review of Main Approaches and Selected Tools. Publications Office of the European Union; European Agency for Safety and Health at Work.
Tucker, S., Diekrager, D., Turner, N., Kelloway, E.K., 2014. Work-related injury underreporting among young workers: prevalence, gender differences, and explanations for underreporting. J. Saf. Res. 50, 67–73.
Turner, N., Tucker, S., Kelloway, E.K., 2015. Prevalence and demographic differences in microaccidents and safety behaviors among young workers in Canada. J. Saf. Res. 53, 39–43.
US Department of Transportation-USDOT, 2011. Safety Culture: A Significant Driver Affecting Safety in Transportation. Research paper prepared for the US DOT Safety Council.
Veltri, A., Pagel, M., Behm, M., Das, A., 2007. A data-based evaluation of the relationship between occupational safety and operating performance. J. SHE Res. 4 (1), 1–22.
Von Eye, A., Schuster, C., 1998. Regression Analysis for Social Sciences. Elsevier.
Wachter, J.K., Ferguson, L.H., 2013. Fatality prevention: findings from the 2012 forum. Profess. Saf. 58 (7), 41–49.
Waltz, C.F., Strickland, O.L., Lenz, E.R., 2005. Measurement in Nursing and Health Research, third ed. Springer, New York.
Wood, M.H., Gyenes, Z., 2015. Lessons learned from corrosion-related accidents in petroleum refineries. Loss Prevent. Bull. 246.
Worthen, B.R., White, K.R., Fan, X., Sudweeks, R.R., 1999. Measurement and Assessment in Schools, second ed. Allyn & Bacon/Longman, New York.
Wreathall, J., 1995. Organizational culture, behavioral norms and safety. In: Carnino, A. (Ed.) (International Atomic Energy Agency, Vienna (Austria)); Weimann, G. (Ed.) (Oesterreichisches Forschungszentrum Seibersdorf GmbH (Austria)); American Nuclear Society, Vienna (Austria). Austria Local Section; Oesterreichisches Forschungszentrum Seibersdorf GmbH (Austria), vol. 889, pp. 75–83 (ISBN 3-9500255-3-7; Worldcat; Apr 1995; p; ANS-ALS; Vienna (Austria); International topical meeting on safety culture in nuclear installations; Vienna (Austria); 24-28 Apr 1995).
Yates, G.R., Bernd, D.L., Sayles, S.M., Stockmeier, C.A., Burke, G., Merti, G.E., 2005. Building and sustaining a system wide culture of safety. Joint Commission J. Qual. Patient Saf. 31 (12), 684–689.
Zeng, S.X., Shi, J.J., Lou, G.X., 2007. A synergetic model for implementing an integrated management system: an empirical study in China. J. Cleaner Prod. 15 (18), 1760–1767.