

TESTING THE VALIDITY AND RELIABILITY OF A SAFETY CULTURE MODEL USING PROCESS AND OCCUPATIONAL SAFETY PERFORMANCE DATA

Ronny Lardner, The Keil Centre Ltd, UK
Paul McCormick and Emily Novatsis, Woodside, Australia

INTRODUCTION

DEVELOPMENT OF OUR SAFETY CULTURE FRAMEWORK

Given the well-established knowledge that a strong safety culture leads to excellent health, safety and integrity performance, organisations have adopted models to define and measure their safety culture. Many such models have arguably been biased towards occupational safety, and have paid insufficient attention to the cultural components of process safety. Moreover, few companies develop their own safety culture model and, importantly, test its validity and reliability. In other words, does the safety culture model measure safety culture effectively? This paper describes the process undertaken by Woodside to establish that the framework it developed to align and improve its process and occupational safety culture is valid and reliable.

Woodside had information from several sources, including recent and repeat incidents, safety climate survey results, and outcomes from an integrity management review all indicating a need for safety culture improvement. In response, Woodside reviewed different methods to improve its safety culture and decided that developing a competency framework for safety behaviour was most appropriate. Such an approach describes safety culture in terms of the behaviours required from managers, supervisors and everyone in the workforce, specifies how these behaviours relate to each other, and readily lends itself to integration into the organisation's safety management and human resource systems. Moreover, this approach provides a common language and understanding of safety culture, whilst enabling flexible application across the organisation.

This method for operationally defining safety culture was pioneered by Wood Group Engineering in conjunction with The Keil Centre. Well grounded in research, inputs to the competency model included academic studies (e.g. HSE, 1999; HSE, 2001; HSE, 2003; Flin, Mearns, O'Connor & Bryden, 2000; Zohar, 2002), industry lessons (Step-Change in Safety, 2004), and internal company research using critical incident interviewing (Flanagan, 1954) and the repertory grid technique (Kelly, 1955). The resulting competency model comprised four behavioural themes – Standards, Communication, Risk Management and Involvement – and was linked via three occupational groups – Managers, Supervisors and Everyone. Specific positive and negative safety behaviours within each theme for each group were identified.

Wood Group Engineering kindly permitted Woodside to develop its own version of the competency model. Supported by The Keil Centre, Woodside conducted its own

internal research where safety climate survey findings, incident review outcomes, interviews with a cross section of the workforce and the high-reliability organization literature (e.g. Weick, Sutcliffe & Obstfeld, 1999) were considered. These inputs revealed the same four behavioural themes across three occupational groups that were defined in the original competency model. However, some of the behaviours and language were modified to accommodate Woodside context, issues, and influences – in particular, lessons from the Esso Longford gas plant explosion (Hopkins, 2000). Woodside calls this model 'Our Safety Culture'. An overview of the framework and an example competency are provided in Appendix A, Figures 1 and 2. For a detailed description of the development of this type of competency framework, see Hayes, Lardner, Medina and Smith (2007), and Hayes, Novatsis and Lardner (2008).

IMPLEMENTATION OF OUR SAFETY CULTURE FRAMEWORK

Woodside began implementing their Our Safety Culture framework in 2007. Sustained and determined use of various methods is promoted to influence behaviour and culture over time. Applications include: gap analysis against the standard of behaviour described in Our Safety Culture, design of training courses and development activities, inductions for new starters, personnel selection assessment methods, objective setting at performance appraisal, incident investigation, integration into technical health and safety standards, reward and recognition, contractor management, showing links to existing tools and practices, applying behavioural toolkits to reinforce behaviours and developing a range of support material to aid reinforcement. For a detailed explanation of the implementation approach see Hayes, Novatsis and Lardner (2008).

THE IMPORTANCE OF TESTING A SAFETY CULTURE FRAMEWORK

Early commitment and support from the organisation determined that Our Safety Culture framework and methods will be a critical ongoing part of Woodside's health and safety improvement approach. Woodside was confident of the model's merits, given it was developed largely from existing external research. However, the decision was taken to conduct internal research to check that the behaviours in the model are related to occupational and process safety performance within Woodside. Checking the integrity of a newly introduced culture framework is a step seldom

taken by organisations. Yet the information can be used to strengthen the conviction that promoting and displaying behaviours in the framework is imperative for health and safety improvement. Moreover, using a valid and reliable measure of safety culture allows the organisation to assess its safety culture and direct interventions for improvement with confidence. To use an engineering analogy, before installing a valve it undergoes a series of tests: Does it look like a valve? Does it have the components of a valve? Does it perform like a valve is expected to perform when in service? Is the leak test of the valve reliable?

METHOD

OVERVIEW OF THE TESTING APPROACH

Our Safety Culture competency framework was tested by administering the framework as a survey to quantitatively measure the safety behaviours. All behaviours were rated on a six-point Likert rating scale, ranging from 1 (never) to 6 (always), where participants were asked to indicate the frequency with which they perceived the behaviours in their facility or area. In the same questionnaire, additional measures or scales were administered as described below. This data was then used to examine the properties of the survey, namely, its validity and reliability using various statistical techniques. These tests help determine the integrity or effectiveness of the questionnaire and thus framework.

VALIDITY

The validity of a questionnaire involves understanding what it measures and how well it does so. There are different types of validity. Convergent, criterion and factorial validity were examined in the present study. Face validity and content validity are also detailed below, although did not form part of the validation research as they were established previously.

Face Validity

Face validity refers to whether the questions appear valid to experienced industry users. This type of validity is a desirable feature of a survey or framework because it will be better accepted if it looks valid to users. Experienced Woodside personnel find Our Safety Culture framework convincing and relevant.

Content Validity

Content validity refers to whether the content of a survey covers a representative sample of the behavioural domain to be measured – health and safety in this case. The content validity of the framework was established by developing the model from existing academic and industry research and has been described elsewhere (Hunter & Lardner, 2007; Hayes, Novatsis & Lardner, 2008).

Convergent Validity

Convergent validity describes whether there is a relationship between scores on the measure of interest and scores on existing validated measures of the same construct, in

this case, safety culture or climate. The extent that the measure ‘converges’ or is associated with the existing measure provides evidence of the new scale’s validity. In this study, scores on Our Safety Culture scale were correlated with Zohar’s (2000) safety climate scale, assessing supervisor commitment to safety, which has been shown to relate to relevant safety outcomes. A positive relationship between scores on Our Safety Culture measure and Zohar’s (2000) safety climate measure will provide evidence of the new scale’s convergent validity.

Criterion Validity

Criterion validity procedures establish whether there is a relationship between scores on a test or measure and scores on relevant performance criteria. In this study, Our Safety Culture scores were associated with self reported experience of three incident types: injury, minor injury and near miss. This data was collected using a format similar to that used by Zacharatos, Barling and Iverson (2005) where respondents indicate the extent and severity of incidents they have experienced. Negative correlations between Our Safety Culture scores and incident involvement scores will indicate that the new measure is appropriately related to those performance criteria. Moreover, the association between Our Safety Culture scores and self reported safety behaviour scores using Neal, Griffin and Hart’s (2000) measures of safety compliance and safety participation will be examined. Positive correlations between Our Safety Culture scores and scores on these measures will provide further evidence of the new scale’s relationship to relevant performance criteria.

Finally, Our Safety Culture scores from six offshore installations were correlated with installation-specific occupational and process safety rates. This analysis establishes whether those installations with a better safety culture, defined as higher average ratings of Our Safety Culture behaviours, have lower occupational and process safety incident rates. These negative correlations will provide further evidence of the new scale’s criterion validity.

Factorial Validity

Factorial validity procedures involve conducting refined statistical techniques to examine the interrelationships of behaviour data. In this study, a technique called confirmatory factor analysis was employed. Put simply, this technique is used to assess the number of ‘factors’ contained in a model – in Our Safety Culture model Standards, Communication, Risk Management and Involvement are the defined factors – and how the specific behaviours in the model ‘load onto’ those factors.

RELIABILITY

The reliability of a survey refers to its consistency. In other words, it is the degree to which the survey measures the same way each time it is used under similar conditions by the same participants. The method used to assess reliability in this study is called internal consistency reliability. The

statistic used to measure the internal consistency reliability is called Cronbach's alpha. Put simply, Cronbach's alpha measures how well a set of items in a survey measure a single concept or construct. For example, the items in the 'Standards' theme of Our Safety Culture should correlate with one another. The closer the Cronbach's alpha is to one, the higher the reliability estimate of the survey or subscale of the survey. Nunnally (1978) recommends that scales should exceed 0.70 to be considered sufficiently reliable.

PROCEDURE

Identifying and Selling the Research Opportunity

Large scale data collection was required to conduct the research. Woodside's Production Division had implemented a health and safety perception survey across seven operating assets and technical integrity support group annually, commencing in 2006. Having identified the means by which the data could be collected for the research, management support was sought and gained.

The most significant obstacle to overcome was potential loss of longitudinal data given that one survey had been completed in 2006 and 2007 and a survey based on Our Safety Culture items was proposed for 2008. This concern was mitigated by requesting that 10 per cent of participants complete both the old and new surveys, enabling mapping of responses and thus comparison of scores between years. Moreover, given that the Production Division proposed to focus health and safety development activities around Our Safety Culture framework, it was logical to begin measuring perception of the behaviours in the framework. Another minor concern was the increased length of the survey due to inclusion of the additional scales to conduct the validation. However, this issue was easily allayed given that the longer survey was a once off event.

Engaging a Research Institution

The decision to engage an academic institution was based on the expertise and credibility it brought to analysis and research of this type. A local Australian institution was selected for practical reasons – regular face-to-face meetings and data collection – as well as to support the development of local academic-industry links. Another successful research project was being undertaken for Woodside by the same institution and organisational psychologists in its Department of Psychology had a good track record of conducting health and safety research. Aside from the research purpose, an independent institution was engaged for ethical reasons; namely, to assure anonymity to survey participants.

Implementing the Survey

Implementation of the survey was led internally by the third author, Woodside's Human Factors Adviser. Necessary communications prior to survey distribution were completed by the Production Division health and safety manager. Surveys were sent to each Production facility to

a nominated focal point who distributed the surveys during pre-shift meetings. Data were collected over approximately six weeks to capture all swings and the technical integrity support group. Completed surveys were returned to the first author via internal mail and subsequently forwarded to the university for analysis. The survey was completed by 648 personnel in 2008 and 1285 personnel in 2009, with response rates of 74% and 66% respectively. The reports on the survey results and research outcomes were compiled jointly by the university representative and Woodside's Human Factors Adviser, who also presented the findings to management. Further dissemination of results was undertaken by Production Division managers.

RESULTS

VALIDITY RESULTS

The results provided solid evidence for the survey's convergent, criterion and factorial validity. It is unusual to obtain such positive validation data for a new survey. This outcome is a reflection of the model being developed from a strong research base.

Table 1 depicts the convergent and criterion validity results for Our Safety Culture survey conducted using the 2008 survey data. The Everyone, Supervisor and Manager subscales of Our Safety Culture framework all showed significant positive correlations with Zohar's (2000) safety climate scale, demonstrating strong evidence of each scale's convergent validity.

Moreover, the Everyone, Supervisor and Manager subscales showed significant positive correlations with Neal, Griffin and Hart's (2000) safety compliance and safety participation scales as well as negative correlations with every self-reported incident measure. These findings are also reported in Table 1 and provide sound evidence of the scale's criterion validity.

With respect to the factorial validity, a four factor correlated model was the best fit for the Everyone, Supervisor and Manager subscales. This outcome indicates that all scales have acceptable factorial validity.

The 2009 analysis of levels of role-specific safety behaviours (managers, supervisors and other employees) demonstrated strong supportive relationships with occupational and process safety incident rates across six different oil and gas installations. This is summarized in Table 2 below.

The technical integrity incident rate was derived by reviewing all incident reports for the six installations for 2009. Relevant incidents were categorised using definitions similar to the CCPS process safety leading and lagging metrics (CCPS, 2008).

A remarkably consistent and strong set of relationships are evident in Table 2, indicating that those facilities with more of the 'right' behaviours by Managers, Supervisors and Everyone tend to have better occupational and process safety performance. These results provide further evidence for the criterion validity of the Our Safety Culture model.

Table 1. Convergent and criterion validity results

Our Safety Culture framework job level	Convergent validity		Criterion validity	
	Correlation with Zohar's (2000) safety climate scale	Correlation with Neal, Griffin & Hart's (2000) safety compliance scale	Correlation with Neal, Griffin & Hart's (2000) safety participation scale	Correlation with number of self-reported incidents
Everyone scale	.434**	.400**	.372**	Near miss – .179** Minor injury – .126** Injury – .202**
Supervisor scale	.721**	.375**	.314**	Near miss – .199** Minor injury – .134** Injury – .164**
Manager scale	.453**	.283**	.231**	Near miss – .132** Minor injury – .122** Injury – .122**

**Correlation is statistically significant at the .01 level.

INTERNAL CONSISTENCY RELIABILITY RESULTS

The internal consistency reliability of the subscales in Our Safety Culture survey was tested using the Cronbach's alpha statistic. Table 3 illustrates that all subscales demonstrated acceptable internal consistency reliability, with all subscales exceeding the .70 value recommended by Nunnally (1978). This result means that the items in Our Safety Culture framework provide a reliable or consistent measure of safety culture.

DISCUSSION

LESSONS LEARNED

An independent academic validation of Woodside's Our Safety Culture framework has added additional credibility to the organisation's safety culture development approach. In a business comprising many people with technical and science backgrounds, using science to prove the merits of

an approach is valued. The findings help strengthen the conviction that promoting, displaying and improving Our Safety Culture behaviours is imperative for process and occupational health and safety improvement. In particular, the results help reinforce to managers and supervisors the critical role their behaviours play in shaping safety outcomes. Woodside has shared these research findings with other companies who are implementing similar health and safety competency frameworks. This information can similarly help these companies to promote the merits of their behaviour frameworks.

Having a human factors specialist in the business with an organisational psychology background was necessary to identify the research opportunity, advise how such research could be conducted and lead this work internally. It was also necessary for the psychologist involved to work with a line manager on presentation of the research results. This approach enabled a balance between academic terminology

Table 2. Criterion validity – Correlations between average Our Safety Culture behaviours and safety performance data across six offshore installations

Our Safety Culture framework job level	Occupational safety performance data			Process safety performance data
	First aid case frequency rate	Total recordable frequency rate	Near miss frequency rate	Technical integrity incident rate
Everyone scale	–0.54	–0.32	–0.64	–0.53
Supervisor scale	–0.56	–0.90	–0.50	–0.73
Manager scale	–0.60	–0.45	–0.80	–0.71

Notes

(1) All correlations rounded to nearest two decimal places

(2) All performance rates were calculated as incidents per million man hours, for each installation

(3) Correlations can be classified as weak (0.1 to 0.2), moderate (0.2 to 0.35) and strong (>0.35)

Table 3. Internal consistency reliability results

Our Safety Culture framework job level	Our Safety Culture framework subscale	Cronbach's alpha
Everyone scale	Standards	.78
	Communication	.89
	Risk Management	.88
	Involvement	.88
Manager scale	Standards	.88
	Communication	.91
	Risk Management	.93
	Involvement	.93
Supervisor scale	Standards	.91
	Communication	.93
	Risk Management	.93
	Involvement	.94

and practical language such that results were communicated in a meaningful way for a broader business audience.

Although not an issue experienced at Woodside, data indicating that manager and supervisor behaviours correlate with process and occupational safety incident data could be confronting and potentially sensitive. For organisations undertaking similar research, it is important to gauge managers' preparedness to hear the results. It may be necessary to inform managers early on about the anticipated relationships and promote how such information can be used to influence and facilitate improvement.

LIMITATIONS OF THE STUDY AND FURTHER RESEARCH

A key limitation is that correlation does not mean causation. The study does not determine whether Our Safety Culture behaviours influence involvement in incidents or whether involvement in incidents influences these behaviours. The

latter option is unlikely though, given it is inconsistent with existing research. Still, to build on this study and overcome this limitation, Woodside has commissioned a longitudinal study to track perceptions of Our Safety Culture behaviours over three years and clarify whether these perceptions have a lagged effect on incident involvement. This study will supplement the existing validation by providing an indication of the framework's predictive validity – that is, examining whether Woodside's measure of safety culture can predict future safety outcomes such as incident involvement.

Another limitation of this study is that some data were self-reported. The concept relevant here is common method bias and suggests that significant correlations between two variables may be due to the common method that the data were collected by, rather than any real relationship (Podsakoff, MacKenzie, Lee & Podsakoff, 2003). This study used scales with different response categories and assured anonymity to reduce social desirability, which would have helped mitigate potential biases. However, in future research, information from the organisation's incident database will be used for the criterion measures.

CONCLUSION

This study summarised how Woodside developed its safety culture framework and described the process undertaken to test its effectiveness. The research confirmed that Woodside's Our Safety Culture framework is valid and reliable, and that installations with better safety culture ratings tended to have better process and occupational safety performance. Future assessments of safety culture using this framework and decisions on where to direct interventions for improvements can be made with confidence. The results have helped strengthen the argument that promoting and displaying Our Safety Culture behaviours is critical to Woodside's continued health and safety improvement effort, and will also prove very useful to other organisations adopting a similar approach.

APPENDIX A

Theme	Everyone	Supervisors	Managers
Standards	Follow rules	Ensure compliance	Set high standards
Communication	Speak up	Encourage the team	Communicate openly
Risk management	Be mindful	Promote risk awareness	Confront risk
Involvement	Get involved	Involve the team	Involve the workforce

Figure 1. Overview of our safety culture framework



Figure 2. Example our safety culture competency: managers, communicate openly

REFERENCES

- CCPS (2008). *Process Safety: Leading and Lagging Metrics*. New York: Centre for Chemical Process Safety.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–358.
- Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Safety climate: Identifying the common features. *Safety Science*, 34, 177–192.
- Hayes, A., Lardner, R., Medina, Z., & Smith, J. (2007). Personalising safety culture: What does it mean for me? *Paper presented at Loss Prevention conference, 22–24 May 2007*. Edinburgh, UK.
- Hayes, A., Novatsis, E., & Lardner, R. (2008). Our Safety Culture: Our behaviour is the key. *Paper presented at Society of Petroleum Engineers HSE Conference, 15–17 April, 2008*. Nice, France.
- Hopkins, A. (2000). *Lessons from Longford: the Esso Gas Plant Explosion*. Sydney: CCH Australia Ltd.
- HSE (1999). *Effective supervisory leadership behaviours in offshore oil and gas industry*. <http://www.hse.gov.uk/research/otopdf/1999/oto99065.pdf>
- HSE (2001). *Involving employees in health and safety: Forming partnerships in the chemical industry*. HSG 217 ASBN 0717620530.
- HSE (2003). *The role of managerial leadership in determining workplace safety outcomes*. <http://www.hse.gov.uk/research/rrpdf/rr044.pdf>
- Hunter, J., & Lardner, R. (2008). Unlocking safety Culture Excellence: Our behaviour is the key. *Paper presented at Institute of Chemical Engineers Hazards Conference, April 2008*. Manchester, UK.
- Kelly, G. (1955). *The Psychology of Personal Constructs*. New York: Norton.
- Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behaviour. *Safety Science*, 34, 99–109.
- Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioural research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.
- Step-Change in Safety (2004). *Fatality Report – Report into offshore fatalities 2000–2002*. Step Change in Safety: www.stepchangeinsafety.net
- Weick, K., Sutcliffe, K., & Obstfeld, D. (1999). Organising for high reliability: processes of collective mindfulness. *Research in Organisational Behaviour*, 21, 81–123.
- Zacharatos, A., Barling, J., & Iverson, R. D. (2005). High-performance work systems and occupational safety. *Journal of Applied Psychology*, 90, 77–93.
- Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on microaccidents in manufacturing jobs. *Journal of Applied Psychology*, 85(4), 587–596.
- Zohar, D. (2002). Modifying supervisory practices to improve sub-unit safety: A leadership-based intervention model. *Journal of Applied Psychology*, 87, 156–163.