Ranking priorities for safety improvement: a study of methods

Anthea Ashford[1], Liz Hellier[2] and Andrew Weyman[3]

[1]Insitute of Naval Medicine, UK, [2]Open University, UK, [3]University of Bath, UK

ABSTRACT

Effective safety management relies on the identification of vulnerabilities. Tapping employee insights is a valuable source of intelligence. Of the array of qualitative and quantitative elicitation techniques, staff surveys are the most commonly encountered: psychometric measures of situational influences on workplace climate provide a valuable means of benchmarking and monitoring. However, they afford little insight about the relative importance of the constituent themes – they are essentially silent on the issue of prioritising topics for intervention. In recognition of this, an arising question relates to determining the strengths and limitations of alternative elicitation techniques. The study reported here compared the performance of three widely used methods of priority elicitation techniques: direct ranking, Q-sort and the method of paired comparisons, for a set of nine features of workplace safety climate. Results showed high agreement with respect to the rank order produced by the three methods. A point of contrast was that the interval scale output from the method of paired comparisons provided an indication of the relative difference (proportional importance) between the variables. However, this method was lengthier to complete. The relative merits of the three methods and the implications for eliciting priorities in safety management are discussed. Areas of future research were identified to develop these measures further for use in the safety management arena.

KEYWORDS

Priority elicitation methods, safety management.

Introduction

Effective safety management relies upon the ability of organisations to implement safety systems, determine their effectiveness and identify vulnerabilities to determine priorities for improvement. Evidence-informed strategic decision making is crucial for allocation of finite resources to those areas that may have the greatest impact in improving safety (Baker et al., 2014). The capacity to tap into employee perspectives is a potentially important component of the evidence base. Safety climate surveys contribute through providing quantifiable ratings of an array of safety management themes (Zohar, 2010). However, they offer little insight into determining employee views on relative priorities for improvement. There are transparent pitfalls to inferring priorities based on themes exhibiting the highest negative ratings, as the respective scales relate to different facets of risk management and types and magnitudes of risk. Alternative complementary methods are needed to elicit employee views on priorities for intervention (Sawhney et al., 2011).

Within the workplace safety and risk management domains, studies have attempted to establish the relative merits and implications of ranking techniques in expert assessments as well as managerial and employee perspectives (see, for example, Comer et al., 1984). However, empirical findings have indicated that the choice of ranking technique may produce notable differences in the ordering of entities and variation in reliability of output, dependent on how the ratings are manipulated and

combined (Comer et al., 1984; van den Fels-Klerx et al., 2017). Therefore, an investigation of the quality and robustness of data generated using ranking techniques requires careful consideration. The study reported in this paper set out to compare the output from three widely used ranking methods when applied to eliciting employee views on priorities for safety climate intervention.

Method

A review of ranking methods was undertaken, with a focus on those utilised in published safety research. The criteria for inclusion were that the methods should: allow identification, assessment and development of priorities; be simple, unambiguous and straightforward to perform; and support the capacity to determine the degree of consensus between respondents. The three most commonly applied techniques that met these criteria were determined to be: the direct ranking method, the Q-sort method and the method of paired comparisons. Direct ranking involves the manipulation of a whole set of stimuli simultaneously to produce a single ordinal (high-low) ranking, with no 'ties'. The Q-sort is an iterative ranking technique, designed to determine the degree of shared perspective over the relative strength or importance of a set of stimuli, resulting in an ordinal scale. The method of paired comparisons requires respondents to indicate which stimuli of two stimuli selected at random from an item set is more important than the other stimuli. This is repeated for all permutations of pairings with the output being an interval scale. The study was a within-subject, repeated measures design with all participants completing each ranking method. Ethical approval to undertake the study was granted by the University of Bath (Reference: Ethics 17-132).

Materials and sample

To test the performance and utility of the methods a set of relevant items to rank was required. An earlier qualitative study (Bennett et al., 2016) and review of the relevant literature revealed several dimensions that were determined to represent a range of sociotechnical organisational aspects that, if present, might indicate positive safety climate features. To reduce the complexity of the exercise, a representative statement for each of the dimensions was used (as advocated by O'Hara et al., 2014). The following nine dimensions and representative statements (here in parentheses) were used: *individual actions* (everyone accepts that flight safety is their responsibility); *management commitment* (individuals are empowered by their management to take actions in the interests of flight safety); *priority of safety* (people do not take flight safety risks, even when work demands are high); *safety training* (flight safety training is an integral part of all routine training); *communication* (occurrences that have flight safety implications are consistently followed up); *safety system* (flight safety risks are considered in the normal planning or briefing cycle); *working environment* (where I work, hazards are appropriately assessed and controlled); *human resources* (there are enough people to do the job safely); *competency/experience* (people here are sufficiently competent and experienced to do the jobs they are required to do safely).

Respondents were required to rank the nine statements according to the bi-polar criterion of 'most like' to 'least like' their workplace. This criterion was chosen as it was determined that respondents could reasonably be expected to observe their workplaces and make a subjective judgement of the degree to which the different dimensions were present.

Respondents were recruited on an opportunity basis from two squadrons of naval aviation personnel, chosen to represent a range of ranks and specialisations. The sample comprised of 37 personnel (15 aircrew and 22 engineers, of which 15 were officers, 11 were senior ratings and 11 were junior ratings).

Procedure

Each method was presented to respondents in the form of a booklet. To minimise order effects, the presentation of methods was counterbalanced using a Latin square design. Four versions of each booklet for each method were made, each with a different presentation order of items (in the case of the method of paired comparisons, the order of pairs presented was randomised). Respondents were briefed, invited to participate and signed informed consent forms. Three data collection sessions were scheduled during the working day, with at least four hours between sessions.

Data analysis

All three methods allowed calculation of between-person concordance, utilising Kendall's coefficient of concordance (*W*). Varying from 0 (no agreement) to 1 (complete agreement), this can be interpreted in a similar way to kappa values. The Friedman's test was used to test the degree of agreement on the level of rank order across the three methods.

Results

There was fair overall agreement for each of the three ranking methods. (Direct ranking method W=0.271, Chi^2 value= 80.2; Q-sort method W= 0.256, Chi^2 value= 75.7; method of paired comparisons W=0.325, Chi^2 value= 96.0), which was statistically significant (p<0.01). The Friedman's test showed that the three ranking methods provided an essentially equivalent rank order ($\chi^2(2) = 1.33$, p=0.513). Given this finding, only the results from the method of paired comparison is detailed below as it allowed the most nuanced insight. In all methods, the statement related to *human resources* was consistently ranked 'least like' respondents' units, so all other items were scaled in reference to this dimension (see Figure 1).

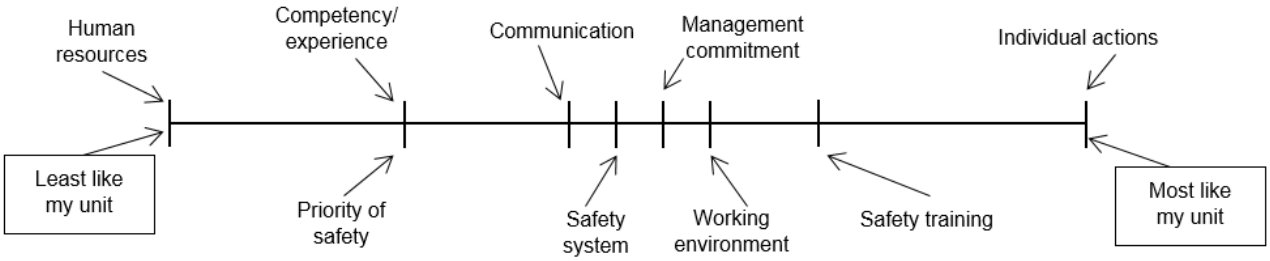

Figure 1: Ratio ranking produced by the method of paired comparisons.

Figure 1 showed that *human resources* was the dimension considered to least reflect the workplace, while *individual actions* most reflected the workplace. *Priority of safety* and *competency/ experience* were ranked more negatively while the remainder of the dimensions showed little discrimination between them.

Discussion

Unlike previous findings, the current study showed that the method of ranking did not significantly appear to affect the aggregated rank order that respondents allocated nine safety climate dimensions. This suggests that all three methods might be suitable for safety managers to use to elicit employee insight into prioritisation. The direct ranking and Q-sort methods were both easy and economical to administer (4-page booklet, approximately 15 minutes to complete) whilst the method of paired comparisons was not (38-page booklet, approximately 30 minutes to complete). However, the method of paired comparisons provided greater insight into the relative priorities, as it gave an indication of the relative distance between the ranked variables, whereas direct ranking and Q-sort produced only ordinal ranks.

Employee insights suggested that safety improvements in these workplaces might best be targeted at (i) ensuring availability of sufficient numbers of personnel, (ii) ensuring sufficient levels of competency/ experience and (iii) focusing on prioritisation of safety. In contrast, it appeared to be well accepted that colleagues in the workplace displayed a high degree of individual responsibility for safety. The remaining climate dimensions appeared to show relatively little discrimination between them. Future research might seek to build on these findings by exploring whether there is agreement as to the rank order of the importance of the dimensions in developing and maintaining a safe working environment. In addition, seeking the views of senior management to indicate whether the views articulated here are mirrored or not, would be a useful avenue for further research. Furthermore, future work could look to explore the relationship between prioritisation through ranking methods and prioritisation through interpretation of results from a safety climate survey.

Conclusions

The three methods of ranking used in this study allowed personnel to make stable and meaningful distinction between the degree of presence of nine safety climate dimensions in the workplace such that they could be used as potential elicitation techniques to identify priorities for improvements. The methods had differing degrees of economy and ease of administration to respondents and further afforded differing levels of insight to interpretation of the results.

References

Baker, R. H. A., Anderson, H., Bishop, S., MacLeod, A., Parkinson, N., Tuffen, M. G. (2014). The UK plant health risk register: A tool for prioritising actions. EPPO Bulletin, 44(2):187-194.

Bennett, A., Hellier, E., Weyman, A. (2016). Characterising influences on safety culture in military aviation: A methodologically grounded approach. In de Waard, D., Brookhuis, K. A., Toffetti, A., Stuiver, A., Weikert, C., Coelho, D., Manzey, D., Ünal, A. B., Röttger, S., Merat, N. (Eds) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfeseurope.org.

Comer, M. K., Seaver, D. A., Stillwell, W. G., Gaddy, C. D. (1984). Generating human reliability estimates using expert judgement. US Nuclear Regulatory Commission. NUREG/CR-3688/SAND84-7115.

O'Hara, R., Johnson, M., Hirst, E., Weyman, A., Shaw, D., Mortimer, P., Newman, C., Storey, M., Turner, J., Mason, S., Quinn, T., Shewan, J., Siriwardena, N. (2014). A qualitative study of decision-making and safety in ambulance service transitions. Health Services and Delivery Research, 2 (56).

Sawhney, G., Cigularov, K. P. Chen, P., DeJoy, D., Huang, Y-H., Kelloway, E. K., Scharf, T. (2011). Safety climate theory, measurement and development: establishing a five-year research agenda. Meeting of National Institute for Occupational Safety and Health and the American Psychological Association, Florida May 2011.

Van der Fels-Klerx, H. J., Van Asselt, E. D., Raley, M., Poulsen, M., Korsgaard, H., Bredsdorff, L., Nauta, M., D'agostino, M., Coles, D., Marvin, H. J. P., Frewer, L. J. (2017). Critical review of methods for risk ranking of food-related hazards, based on risks for human health. Critical Reviews in Food Science and Nutrition, 1-16.

Zohar, D. (2010). Thirty years of safety climate research: reflections and future directions. Accident Analysis and Prevention, 42:1517-1522.