The Accident Risk Management Questionnaire (ARM-Q): A Report on Two Validation Studies

Gerard J. Fogarty (fogarty@usq.edu.au)
University of Southern Queensland, Toowoomba QLD 4350 Australia

Todd Shardlow (todd.shardlow@qr.com.au)
Queensland Rail, Brisbane QLD 4000 Australia

Abstract

Safety is a major concern in the modern workplace and it is now not uncommon to see safety attitude questionnaires used in selection, as well as training settings. As with any other psychometric instrument, it is essential that the usage of such tests be justified through validation studies. This paper reports on two validation studies of the Accident Risk Management Questionnaire (ARM-Q), an instrument that is widely used for selection and training purposes in Australia and overseas. The first study analysed the profiles of 159 transport workers and found tentative evidence for the factorial validity of the ARM-Q. The second study examined the predictive validity of the questionnaire. ARM-Q scores were collected from 680 employees working for 14 different road transport companies in Australia. In the months that followed, data were also collected on safety measures such as accident rates, workers' compensation costs, and hours lost through accidents. Meta-analytic techniques were then used to estimate the relationship between scores on the ARM-Q and these safety indicators. The results showed small but reliable relationships, supporting the test developer's claims that the instrument has predictive validity in work settings such as those found in the present study.

Introduction

Workplace accidents and injuries are the scourge of both employees and companies alike. The financial cost of these accidents is incurred through damage to property and infrastructure, medical costs, increased insurance premiums, disability or workers' compensation payments, absenteeism, or by lost revenue due to decreased production. It is generally accepted amongst researchers that human error or unsafe human behaviour is associated with 80-90% of accidents (Lawton & Parker, 1998; Reason, 1997). It is therefore important that we develop our knowledge of the human antecedents to accidents so that interventions can be designed to reduce their occurrence and subsequent impact on the lives and fortunes of individuals and companies.

Whilst there are some broad organizational factors that contribute in a major way to workplace safety, the present study will focus on individual traits and attitudes that shape safety behaviours. Specifically, it will focus on a small number of factors that are captured by a test instrument that has been developed to identify those people at higher risk for workplace accidents. The twin aims of the study are to test the predictive validity of the instrument in a road transport setting and, at the same time, to explore the role of individual differences in safe work performance. We begin by reviewing the literature on these factors before proceeding to a description of the instrument and a report of two validation studies.

Safety locus of control. The concept of locus of control refers to the degree to which individuals perceive that consequences of their actions and other life events are controlled by their personal effort. In the context of safe work performance, individuals with an internal safety locus of control feel in control of the outcomes of their behaviour, and take responsibility for job safety and any injuries that they may or may not incur (Lawton & Parker, 1998). Conversely, individuals with a high external safety locus of control exhibit low safety consciousness and do not see a cause and effect relationship between their actions and safety outcomes. Safety outcomes are viewed as being beyond their control and blame is attributed to external factors such as other people's actions, chance events, or bad luck (Lawton & Parker).

Stress Tolerance. Stress is a variable that should be considered in any aspect of work performance. Its effects can be seen everywhere. Hoffman and Stetzer (1996) found that role overload led to employees taking shortcuts in order to get work completed. Fogarty (2004) showed that high stress levels were associated with higher error rates among aircraft maintenance engineers. Lawton and Parker (1998) suggested that stress acted as a mediator between personality variables and accident rates. However, recent models of stress (e.g., Hart & Cooper, 2001) remind researchers that it is important to focus on the positive as well as the negative aspects of the work environment. If susceptibility to stress leads to accidents (Lawton & Parker, 1998), it follows that individuals with a high tolerance for stress should therefore be less susceptible to accidents.

Risk Avoidance. In their architecture of employee attitudes to safety, Cox and Cox (1991) described risk as one of the major aspects of safety climate, reflecting both the safeness of the work environment and the idea of personal immunity. Likewise, in a review of safety climate research focusing on risk, Dedobbeleer and Béland (1998) concluded that risk perception is an

important dimension of safety climate and closely linked with employee commitment to safety standards. Similarly, Morrow and Crum (1998) looked at the importance of safety perceptions by comparing perceived and objective measures of safety risk on employee outcomes among railroad employees. They found that subjective perceptions of work safety were actually a better predictor of employee outcomes - such as job involvement, stress, organisational commitment - than objective measures of risk and safety in the organisation.

Driver Attitude. It is not uncommon to find context-specific measures of safety attitudes but they are often also subsumed under other factors such as risk, stress, and sensation seeking. Trimpop, Austin, and Kirkcaldy (2000), for example, found that traffic accidents were best predicted by attitudes to traffic risk, stress, and driver distance (exposure). The levels of stress experienced and hours worked were the best predictors of other work-related accidents, with risk attitude and work-related driving distance also predicting accidents to a lesser extent.

Test Description

The Employee Safety Inventory (ESI) was developed by NCS/London House to measure safety awareness and attitudes towards safety behaviours. An Australian/New Zealand adaptation and validation was undertaken by People and Quality Solutions (PaQS) in 1992-93. Since then, the local version has been known as the Accident Risk Mangement Questionnaire (ARM-Q). The ARM-Q is a not an industry-specific questionnaire. It is currently used for selection and training purposes and administered approximately 5,000 times per year.

The ARM-Q consists of 117 items and comprises eight subscales: Safety Control, Risk Avoidance, Stress Tolerance, Driver Attitude, Quality Attitude, Safety Index, Accuracy, and Distortion. All but one item utilise a six point Likert-type scale, with responses ranging from a score of one indicating "Strongly Agree", to a score of six indicating "Strongly Disagree".

Safety Control is the first subscale measured by the ARMQ. It comprises 26 items and pertains to an individual's locus of control in relation to safety and accident prevention. Individuals who possess a high internal locus of control take responsibility for any accidents that occur as being a result of their own actions and also assume personal responsibility for job safety and accident prevention. A low Safety Control score indicates that the individual possesses an external locus of control and tends to believe external factors such as other peoples' actions, luck, or fate are the cause of accidents that occur. An example of an item on this scale is: *With my luck, I will probably have an accident in the near future.*

There are 24 items that contribute to the Risk Avoidance subscale. This scale is a measure of the individual's ability to perceive a safety risk. It also examines the propensity of the individual to engage in behaviours that may be dangerous or counter-productive, such as not wearing personal protective equipment that the company has provided. An example of an item on this scale is: *I occasionally like to do my own thing at work instead of always following company rules.*

The Stress Tolerance subscale contains 39 items, and examines the individual's on-going experience with stress and ability to cope with stress, as opposed to measuring the degree of total stress in an individual's life. Stressed employees are potentially at a higher risk of being involved in accidents due to increased fatigue and an increased susceptibility to distraction on the job. An example of an item on this scale is: *Maybe I overreact sometimes, but little things at work can really get on my nerves.*

The Driver Attitude subscale is made up of six items, and assesses the individual's attitudes towards safe driving practices and serves as an indication of increased likelihood of being involved in a motor vehicle accident. An example of an item on this scale is: *Drivers can prevent nearly all motor vehicle accidents by being careful and following all traffic regulations.*

Quality Attitude is the next subscale on the ARM-Q, and consists of 18 items assessing the individual's attitude to the overall quality of the work being performed, including the quality of the individual's personal work habits and adherence to company safety procedures. It is also a measure of how committed the individual is to detecting and avoiding errors, in addition to the commitment shown to continually improving the overall quality of the service or product being provided. An example of an item on this scale is: *I am very skilled at quickly spotting, then correcting, any defects in the projects I am working on.*

The Safety Index scale is a composite score generated to provide an overall indication on an individual's work safety attitudes. It is a combination of the individual's scores on the Safety Control, Risk Avoidance, and Stress Tolerance scales.

The ARM-Q also yields two validity scales, Accuracy and Distortion. The Accuracy scale measures how accurately the individual has filled in the questionnaire, and can be used to determine if respondents have literacy or comprehension problems, uncooperative people who respond randomly, or people who may have become distracted whilst filling in the questionnaire. The Distortion scale measures whether individuals have attempted to distort their answers to create a favourable impression of their safety awareness. The last three scales were not evaluated in this study.

Existing Validation Studies

The publishing company released to the authors the abstracts of 26 technical reports on the ESI examining various aspects of reliability and validity of the ESI.

These abstracts indicate that the ESI scales demonstrate sound internal consistency and test-retest reliabilities (generally above .80) and that they predict and covary with relevant criterion behaviours, such as accidents, drug usage, and supervisor ratings of work performance. However, the test validity information is limited by virtue of the fact that it is gleaned from abstracts. In one of the very few published validation studies, Boye, Slora, and Britton (1990) reported that test-retest reliability for the various scales ranged from .75 to .91.

In a more extensive (but unpublished) report prepared for PaQS, Kendall and Want (2001) reviewed technical abstracts supplied by NCS/London House and PaQS. They also analysed 5,000 ARM-Q forms collected between 1994 and 2000 from Australian companies across a variety of industries. Their report supported claims about the factorial validity of the ARM-Q. It also supported positive claims made by London House and PaQS about the criterion validity of the ARM-Q scales and their reliability. Kendall and Want concluded that the ARM-Q meets acceptable professional standards for psychometric instruments employed for selection purposes. Once again, however, the report takes the form of a brief summary of major findings. To date, with the exception of Boye et al. (1990), there are no published validation studies of either the ESI or its Australian equivalent, the ARM-Q.

Comment on the Scope of the Present Validation

As mentioned above, test validation is a lengthy and exhaustive process because of the wide range of aspects that need to be covered. The technical abstracts supplied by London House and PaQS show that many of these aspects have been covered. Our goal in this paper is to move just a small part of the validation process to the public domain. In so doing, we will also seek to compare our findings with those reported in the technical abstracts.

Study 1

Because there are no factor analytic studies in the technical reports, we began by conducting a small construct validation study using 159 ARM-Q profiles supplied by PaQS. The sample came from drivers working for transport companies. The ratio of cases to variables (159 : 117) did not allow a rigorous test of the factor structure of the ARM-Q, so our report on this study is brief and our conclusions somewhat tentative.

Using principal axis factoring with oblique rotation and following a strategy whereby each scale was factor analysed separately (i.e., no attempt was made to factor analyse all 117 items simultaneously), we were able to establish the unidimensionality of all scales except Driver Attitude, where scree plots indicated that two factors should be extracted, one tapping attitudes to safe driving, the other seemingly more related to the drivers'

perceptions of who is to blame for accidents (and probably tapping locus of control).

With the exception of the Driver Attitude scale (α = .73), the Cronbach alpha coefficients for all scales were satisfactory and similar to those reported by the test producers. The lower internal consistency reliability estimate for the Driver Attitude scale can probably be attributed to the multidimensional nature of the scale. We note that its reliability coefficient is also lower in the technical report by Huff and Brasher (1999; α = .75). Scale intercorrelations are shown in Table 1.

Table 1: Correlations among safety scales (N = 159)

	1	2	3	4	5
1. Safety control	**.81**	.67	.63	.58	.71
2. Risk avoidance	.65	**.87**	.63	.35	.60
3. Stress tolerance	.58	.70	**.88**	.33	.51
4. Driver attitude	.44	.34	.27	**.62**	.52
5. Quality attitude	.69	.69	.71	.35	**.80**

Notes.
1. All correlations significant at .001 level.
2. Reliabilities (α) shown in main diagonal.
3. Data from Huff & Brasher (1999) are in the upper triangle (N = 109).

We can see from Table 1 that the scales are significantly correlated but not collinear, a finding consistent with the claim that they tap different underlying constructs.. We can also see that these correlations are similar to those reported by Huff and Brasher (1999), the only London House technical report to publish the scale intercorrelations.

On the basis of these analyses, we conclude that whilst a larger study is needed to establish the factorial validity of the ARM-Q, the existing scales have reasonable internal consistency and do not overlap excessively.

Study 2

The second study forms the primary focus of this paper and seeks to test the predictive validity of the ARM-Q in a road transport setting.

Participants

Participants for the study were 680 employees (mostly male) involved in road transport activities from 14 separate Australian companies. All of these companies had engaged People & Quality Solutions Pty Ltd (PaQS) to conduct an accident risk management assessment for their particular company.

For privacy reasons, the researchers did not have access to information relating to demographic characteristics of the respondents, selection methods for participation in the survey, or the positions of respondents in their various companies. Given that these factors are not taken into consideration in the scoring of ARM-Q

profiles, this information was not considered to be of particular importance for the purposes of this validation study.

Materials

The ARM-Q was used to measure the various dimensions of safety climate. Data provided to researchers were in the form of the individual ARM-Q profiles generated by NCS/London House. Each profile comprised an individual's standardised scores on each of the five major subscales, in addition to scores for Distortion, Accuracy, and a Safety Index composite score.

Information relating to actual safety outcomes was supplied to PaQS by each individual company. This information took the form of a number of different measures of safety outcomes for each individual, including the number of accidents, cost of accidents, absenteeism (hours off work on sick leave), time off work on workers' compensation, and cost of workers' compensation claims.

Procedure

Participants completed the ARM-Q when they attended safety training sessions conducted by PaQS. AMR-Q forms were collected by PaQS and then sent to London House for scoring. Data on the safety performance variables were subsequently collected by each of the 14 companies and forwarded to PaQS for use in follow-up training sessions. The means by which this information was collected by each company varied greatly. Some companies provided information on the basis of in-house records of accidents and their cost, whereas other companies cited information from insurance companies regarding the number and cost of accidents.

Results

A total of 36 cases were excluded by the researchers on the basis of their high scores on the Distortion scale (> 20) or low scores on Accuracy (< 10), leaving 644 cases for analysis.

Following this initial screening, correlations were calculated for all five safety scales. The results are shown in Table 2.

Table 2: Correlations among safety scales (N = 644)

	1	2	3	4	5
1. Safety control					
2. Risk avoidance	.59				
3. Stress tolerance	.52	.52			
4. Driver attitude	.65	.39	.37		
5. Quality attitude	.63	.46	.40	.46	

Note. All correlations significant at .001 level

As was the case in Study 1, the scales of the ARM-Q scales were moderately correlated ($p < .001$).

Given the variability in methodology employed by each company in collecting accident data, it was considered advisable to treat each company as constituting a separate dataset and to use the powerful statistical technique of meta-analysis (Hunter & Schmidt, 1990) in combination with standard multiple regression analysis to assess the strength of the relationship between scale scores and accident statistics. The particular variant of meta-analysis used in this study is known as "bare bones meta-analysis" (Schmidt & Hunter, 2003) because it corrects only for the distorting effects of sampling error without taking other effects (such as attenuation) into account. R^2 values (non-adjusted) were taken as estimates of effect size and were weighted by multiplying each R^2 value by the sample size of that company, before obtaining a weighted average effect size. Table 3 shows the resulting effects across the various performance indicators.

Table 3: Weighted R^2 for different outcome variables

		Safety Outcome Measures				
Company	N	A	B	C	D	E
1	37	.35		.38	.38	.18
2	57	.08	.07			
3	45	.15	.09	.07		
4	23	.55	.44	.38	.44	
5	24		.39		.45	.32
6	62	.09		.13		.06
7	39	.06				.20
8	13	.25	.20			
9	43	.05		.07		.06
10	119			.04	.04	.02
11	42	.18		.07		.34
12	46			.05		
13	49	.13				
14	45	.09				
Mean Weighted R^2		.16	.15	.11	.19	.12

Key:
N = Number of Employees
A = Number of Accidents
B = Cost of Accidents
C = Time off work on Workers' Compensation
D = Cost of Workers' Compensation
E = Hours of Sick Leave

The meta-analyses indicated that approximately 16% of the total variability in number of accidents was predicted by respondents' scores on the ARM-Q. Similarly, respondents' scores accounted for approximately 15% of the total variability in the cost of accidents, 11% of the variance in workers' compensation time, 19% of the variance in the cost of workers' compensation, and 12% of the variability associated with the number of sick leave hours.

These effect sizes are slightly larger than those reported in various London House technical abstracts, which were all conducted on single samples across a range of industries.

Discussion and Conclusions

These findings indicate that the ARM-Q is able to explain significant amounts of variance in safety outcomes (between 11 and 19%). Boye, Joy, Slora, and Jones (1990) reported a similar figure and claimed that when the ESI is used for selection and training purposes, the return on investment (ROI) is likely to be in the region of 1400%. We cannot substantiate the ROI claim but our findings lend support to the various technical reports that show a link between ARM-Q scores and performance. In so doing, they also advance the claims of organizational psychologists that self-report data represent more than just "soft" indicators of how employees feel about their work (c.f., Zohar, 2002). Valid and reliable attitudinal scales can be used to develop appropriate intervention strategies to decrease workplace accidents. These interventions will, in turn, translate into large savings for organizations and better outcomes for individual workers.

In noting the limitations of this study, we point out that we were not able to assess these reliabilities of the dependent variables and so could not correct for attenuation but we were of the impression that the processes for collecting accident data among these 14 companies fell short of the scientific rigour that would be required to ensure high reliability. The true relationship between these attitudinal variables and safety performance is therefore likely to be higher than we have demonstrated here. We also caution that the first study needs to be repeated with a much larger dataset so that the full set of ARM-Q items can be factor-analysed simultaneously. Because we were not able to do this, our findings regarding the factorial validity of the ARM-Q have to be regarded as tentative.

In conclusion, this paper has addressed both theoretical and practical concerns. From a theoretical point of view, we were able to gather attitudinal data and link them with subsequent objective measures of safety performance. In so doing, we have been able to address the need for additional empirical evidence linking attitudinal data with objectively measured work performance outcomes. From a practical point of view, the study contributes to the lengthy validation process to which all selection tests should be subjected and places the findings in the public domain.

References

Boye, M. W., Slora, K. B., & Britton, C. S. (1990). Temporal stability of a safety attitude inventory. *Psychological Reports, 67*, 467-470.

Boye, M. W., Joy, D. S., Slora, K. B., & Jones, J. W. (1990). The relation of the Employee Safety Inventory to accidents and related costs at a national trucking company. *ESI Research Abstract No.13*. Rosemount, Il: London House.

Boye, M. W., Slora, K. B., & Jones, J. (1989). Reliability of the Employee Safety Inventory. *ESI Research Abstract No.7*. Rosemount, Il: London House.

Cox, S., & Cox, T. (1991). The structure of employee attitudes to safety: A European example. *Work & Stress, 5*(2), 93-106.

Dedobbeleer, N., & Béland, F. (1998). Is risk perception one of the dimensions of safety climate? In A. Feyer & A. Williamson (Eds.), *Occupational injury: Risk prevention and intervention* (pp. 73-81). London: Taylor & Francis.

Hart, P., & Cooper, C. L. (2001). Occupational stress: Toward a more integrated framework. In N. Anderson & D. S. Ones & H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of industrial, work, and organisational psychology (vol 2: Personnel Psychology)* (Vol. 2, pp. 93-114). London: Sage.

Hofmann, D., & Stetzer, A. (1996). A cross-level investigation of factors influencing unsafe behaviours and accidents. *Personnel Psychology, 49*(2), 307-339.

Huff, J., & Brasher, E. (1999). Validation of the Employee Safety Inventory (ESI-Q[TM]) at a manufacturing company. *ESI Research Abstract No.26*. Rosemount, Il: London House.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: SAGE.

Kendall, I. M., & Want, R. M. (2001). *Independent Accident Risk Management (ARM) analysis and validation opinion*. Retrieved July 31, 2003, from http://www.paqs.com/armq_validation.html

Lawton, R., & Parker, D. (1998). Individual differences in accident liability: A review and integrative approach. *Human Factors, 40*(4), 655-671.

Morrow, P. C., & Crum, M. R. (1998). The effects of perceived and objective safety risk on employee outcomes. *Journal of Vocational Behavior, 53*, 300-313.

Reason, J. (1997). *Managing the risks of organizational accidents*. Aldershot, England: Ashgate Publishing Ltd.

Schmidt, F. L., & Hunter, J. E. (2003). Meta-analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology* (Vol. 2, pp. 533-554). New Jersey: John Wiley & Sons.

Trimpop, R., Austin, E. J., & Kirkcaldy, B. D. (2000). Occupational and traffic accidents among veterinary surgeons. *Stress Medicine, 16*, 243-257.

Zohar, D. (2002). The effects of leadership dimensions, safety climate, and assigned priorities on minor injuries in work groups. *Journal of Organizational Behavior, 23*, 75-92.