



# Development of safety incident coding systems through improving coding reliability



Nikki S. Olsen<sup>\*</sup>, Ann M. Williamson

School of Aviation, The University of New South Wales, Kensington, Sydney, NSW 2052, Australia

## ARTICLE INFO

### Article history:

Received 3 September 2014

Accepted 27 April 2015

Available online 21 May 2015

### Keywords:

Incident classification

Reliability

Safety management system

## ABSTRACT

This paper reviews classification theory sources to develop five research questions concerning factors associated with incident coding system development and use and how these factors affect coding reliability. Firstly, a method was developed to enable the comparison of reliability results obtained using different methods. Second, a statistical and qualitative review of reliability studies was conducted to investigate the influence of the identified factors on the reliability of incident coding systems. As a result several factors were found to have a statistically significant effect on reliability. Four recommendations for system development and use are provided to assist researchers in improving the reliability of incident coding systems in high hazard industries.

© 2015 Elsevier Ltd and The Ergonomics Society. All rights reserved.

## 1. Introduction

### 1.1. Incident coding systems and their testing

Incident and accident investigation is now an integral part of safety management systems in safety critical industries. Identifying how incidents and accidents occur can reveal factors and causes contributing to past incidents that will help target preventive action to reduce the likelihood of future accidents. An increasing number of accident and incident classification and coding systems have been developed for this purpose in recent years. These coding systems aim to provide a standard, systematic framework for identifying the factors that contribute to accident and incident occurrence. Thus they are tools for managing and understanding the often large amount of information gathered about accidents and incidents and for analysing common causal factors across different incidents and over time. The systems vary considerably in the nature of coding used, with some based on simple lists of possible contributing factors (eg. af Wählberg, 2002), others on formal models of accident causation (eg. TRACer, Shorrock and Kirwan, 2002) and some include more in-depth coding of aspects such as different taxonomies of error involvement (eg. HFACS, Wiegmann and Shappell, 2003). Systems also vary in the type and number of people who do the coding ('coders').

Reliability is an essential attribute of all standard and systematic coding systems (Kirwan, 1996; Ross et al., 2004; Stanton and Stevenage, 1998). It is essential that the coding system produces consistent coding of the data regardless of when or by whom the coding is conducted. This includes intercoder reliability, or how well two or more coders agree with one another, and intracoder reliability, or how well a single coder agrees with themselves when coding the same incident on different occasions. Clearly, if coders cannot agree, then there will be variability in the range of codes applied, so making unclear how the incident actually occurred. Without acceptable levels of reliability there is little chance that the classification and coding of incidents will reflect accurately how they occurred. Classification and coding systems with poor reliability will contribute little to the understanding of safety failures in the settings in which they are applied so making them unlikely to contribute to improving safety.

A range of different methods have been used to estimate the reliability of classification and coding systems (Olsen, 2013). High levels of agreement are needed for a classification and coding system to be judged as reliable but many of these systems, even some in common use, have poor or unknown reliabilities (eg. HFACS-ADF, TAPS, HFACS-ME). Better and more reliable tools are needed to code and classify safety incidents, but the best way of going about this is not clear. A few studies have discussed some of the factors that could play a role in limiting the reliability of classifying and coding safety information (Baker and Krokos, 2007; Olsen and Shorrock, 2010; Li and Harris, 2005), but there has been no systematic analysis of what factors play a role or any

<sup>\*</sup> Corresponding author. Tel.: +61 3 51467916.

E-mail address: [Nikki.Olsen@defence.gov.au](mailto:Nikki.Olsen@defence.gov.au) (N.S. Olsen).

empirical study of their influence. The objective of this research was to identify the factors that might influence reliability, and examine the nature of their influence in order to provide guidance to incident coding system developers, users and researchers in safety critical industries.

## 1.2. Reliability of incident classification and coding systems

Existing studies of reliability and coding system development provide evidence to support hypotheses concerning the factors that might influence the reliability of incident classification and coding systems. Firstly, the scope of the system may influence the reliability of the system for various users. Taib et al. (2012) found that a domain-specific medication error taxonomy (such as one designed only to classify medication errors) was more reliable than a generic medication error taxonomy (one that can be used for errors across all medical departments) when used by medical personnel including pharmacists and nurses. These authors suggested that domain-specific systems, where the scope and wording are very similar to that which would be used in the reports generated by individual medical departments, may produce higher reliability regardless of the profession of the coder. This is because they may require less interpretation in matching report terminology to the available system codes. On the other hand, translating detailed medical terminology from individual departments into more generic medical terminology might result in different interpretations by different coders and lead to lower levels of reliability.

Second, the profession or expertise of the coder is also a potential factor that may influence the reliability of incident coding systems. For example, Taib et al. (2012) found that comparing the reliability of coded reports of medication errors, pharmacists were more reliable than nurses regardless of whether they were using domain-specific or generic systems however did not venture to explain why this might be so. Also supporting the potential influence of coder expertise, Krippendorf (2013) argued that reliability will be poor where coders without knowledge of the subject matter are asked to interpret texts with unfamiliar terms, therefore coders should have an appropriate amount of expertise in the field in order to produce high reliability. This is supported by Baker and Krokos (2007), O'Connor and Walker (2011) and Olsen and Shorrock (2010) who raised concerns for the reliability and validity of coding of human error concepts by coders without human factors expertise who may not be able to distinguish between contextual and cognitive concepts.

Third, coder expertise with the specific coding system may also influence reliability. Baysari et al. (2011) argued that a coder's experience with a particular incident coding system may increase reliability as coders will be familiar with how to use it and the codes available. However this view has been challenged by Olsen and Shorrock (2010) who suggested that experience may only serve to establish users' preferred codes which would look like good reliability, but only really reflect a truncation of the range of codes used. Peter and Lauf (2002) similarly argued that experienced coders develop routines over their career that are inconsistent with the original coding methodology or change their understanding of category meanings. If these effects varied from person to person it might be expected that reliability would be lower with experience. Clearly the effects of coder experience need further evaluation.

Fourth, the size of the coding system in terms of the range of coding choices to be made may influence system reliability. A number of researchers have expressed concerns over the appropriate size of incident coding systems (af Wählberg, 2002; Baysari et al., 2011; Jacobs et al., 2007; O'Connor, 2008). Size, here, refers to either the number of coding choices available at each level of the system (when depicted on paper this would be the 'width' of the

system) or the number of levels within the system (when depicted on paper this would be the 'height' of the system) or both. Systems with greater than two levels of analysis tend to involve more specific codes for describing incident phenomena. Some researchers (Makeham et al., 2008; O'Connor, 2008; Olsen and Shorrock, 2010) suggest that as the levels of analysis increase, the reliability decreases. Isaac et al. (2003a) relate the reason for decreasing reliability to the degree of choice offered to the coder – as the coder progresses through each level of the system, more options are available and more choices are made by coders so reducing reliability. Similarly, when considering the 'width' of the system Krippendorf (2013) more explicitly stated that there should be no more than seven categories at each point in the system because coders find it difficult to consider large numbers of categories. Such large numbers of categories are also thought to encourage coders to form coding habits and preferences which has been suggested to artificially inflate reliability, as described above.

There is little consistency in the literature to define what researchers and developers describe as a 'large' or 'small' system. However O'Connor (2008) and Olsen and Shorrock (2010) suggest that DoD-HFACS and HFACS-ADF respectively were too large for reliable coding with the total number of codes for each of these systems 147 and 155 respectively. This is compared to the original version of HFACS that consists of 19 codes. This comparison gives a little appreciation of the description of 'large' and 'small' incident coding systems in the current literature.

Fifth, the terms and concepts used in the system may also influence its reliability. A number of researchers (Chang et al., 2005; Runciman et al., 2009; Shorrock and Kirwan, 2002) argue that terms within systems need to be unambiguous and incorporate both theoretical concepts as well as generally accepted vocabulary in order that researchers can understand each other's work and thus facilitate the systematic collection, aggregation and analysis of relevant information. Ercan et al. (2007) point out that certain terms may invite bias on behalf of users. They identify two types of bias in the terminology of incident coding systems that may affect reliability. The first, loaded terms, refer to terms that cause some meaning or feeling in a defined subject that may affect whether the term is selected or not. For example, the code 'failed to prioritise attention' is more loaded due to the 'failure' term than the code 'distraction' which is more neutral. The second form of bias refers to category names that describe illegal attitudes, behaviours or private subjects and that may cause an unwillingness to select them due to the instigation of potential repercussions. This has been highlighted in the development of the Canadian version of the Human Factors Analysis and Classification System (HFACS). Developers found that air traffic controllers tended to balk at the term 'violations' so they replaced the word with the term 'contraventions' which they found to be more palatable to the users and resulted in the code being used more often (Wiegmann and Shappell, 2003). The use of terms that may cause bias in coders may result in low levels of reliability.

Li and Harris (2005) also suggested that coders were less inclined to use categories where the terms represented abstract concepts and were more inclined to use categories for which more tangible evidence was available from the accident report narratives. For example, they found that categories showing the lowest levels of reliability were those inferring a lack of 'situational awareness' or identifying 'organisational climate' as a contributing factor. It may be that physical concepts such as 'equipment failure' are coded more reliably than more abstract concepts such as 'decision making' because equipment failure may be readily seen whereas the making of a decision is an internal cognitive event. As many incident coding techniques are based on conceptualisations of cognitive failures that are inferred rather than seen it is possible that the

Download English Version:

<https://daneshyari.com/en/article/550960>

Download Persian Version:

<https://daneshyari.com/article/550960>

[Daneshyari.com](https://daneshyari.com)