Title	A methodology for identifying human error in U.S. Navy diving accidents
Author(s)	O'Connor, Paul
Publication Date	2007
Publication Information	O'Connor, P., O'Dea, A., & Melton, J. (2007) 'A methodology for identifying human error in U.S. Navy diving accidents'. Human Factors, 49 (2):214-226.
Link to publisher's version	http://hfs.sagepub.com/content/49/2/214.short
Item record	http://hdl.handle.net/10379/2803

Cite as: O'Connor, P., O'Dea, A., & Melton, J. (2007). A methodology for identifying human error in U.S. Navy diving accidents, *Human Factors, 49(2)*, 214-226.

A METHODOLOGY FOR IDENTIFYING HUMAN ERROR IN U.S. NAVY DIVING ACCIDENTS

ABSTRACT

OBJECTIVES: To better understand how human error contributes to U.S. Navy diving accidents.

BACKGROUND: An analysis of 263 U.S. Navy diving accident and mishap reports revealed that the human factors classification were not informative for further analysis, and 70% of mishaps were attributed to unknown causes; only 23% were attributed to human factors.

METHOD: Five diving fatality reports were examined using the consensual qualitative research (CQR) method to develop a taxonomy of six categories and 21 subcategories for classifying human errors in diving. In addition, 15 critical incident technique (CIT) interviews were conducted with U.S Navy divers who had been involved in a diving accident or near-miss and analyzed using the dive team error taxonomy.

RESULTS: Overall, failures in situation awareness and leadership were the most common human errors made by the dive team.

CONCLUSION: The dive team human error taxonomy could aid in accident investigation, and the training and evaluation of U.S Navy divers.

APPLICATION: The development of the dive team human error taxonomy has generated a number of considerations that researchers should take into account when developing, or adapting, an error taxonomy from one industry to another.

Introduction

Early concepts of industrial accidents were founded on a mechanistic view of organizations. Theories of accident causation were based in a belief that the main causes of accidents were technological. Basic principles of human cognition, motivation, and attention were not considered in the development of preventative measures, a failure resulting in designs that increased the likelihood of human error (Bainbridge, 1987; Norman, 1988). Not until the late 1970s did the focus of accident investigations begin to broaden, and for the first time accidents were viewed as multicausal events implicating both human and technical contributions to those results.

Turner (1978) identified a number of stages in the lead-up to an accident. Reason (1990), building on Turner's work, distinguished between "active" and "latent" failures. Active failures are usually committed by workers at the 'sharp end' and are likely to act as precursors to accidents. Latent failures, on the other hand, are likely to be committed by personnel higher in the organizational hierarchy.

This study aims to identify the causes of U.S Navy diving accidents, with particular focus on those mistakes Reason categorized as active failures. The reason for focusing on active failure is that the types of human errors made by dive team members need to be quantified to improve diving safety and productivity.

To achieve this aim, the authors first examined data collected by the U.S. Navy following diving mishaps. However, this data set did not provide sufficient information about how active failures by dive team members had contributed to accidents. Therefore, two more data sources — diving fatality reports and interviews with divers about accidents or near misses in which they had been involved — were scrutinized, and a taxonomy for categorizing the human errors made by members of the dive team, and based on the fatality reports was developed. The taxonomy was then used to code the data collected from the interviews.

U.S. Navy Diving

The size of a U.S. Navy diving team may vary with operational requirements such as depth, type of equipment to be used, and number of divers required to complete the mission (Commander, Naval Sea Systems Command, 1999). A typical diving team may include the following personnel:

- Diving Officer: He/she is responsible for the safe conduct of all diving operations within a command and is responsible to the Commanding Officer.
- Master Diver: As the most qualified and experienced person to supervise dives, he/she is in charge of overall diving operations and is responsible to the Diving Officer.
- Diving Supervisor: Generally a first-class (senior enlisted) diver with advanced training, he/she is in charge of the actual diving operations for a particular dive or dive series.
- Diving personnel: These divers carry out a dive, tend divers, act as standby/safety divers, log a dive, and communicate with divers.
- Diving Medical Officer: As a medical doctor with training in hyperbaric medicine, he/she provides on-site medical care for divers.

Although U.S. Navy diving is remarkably safe, because of the high-risk environment in which the divers work, accidents do occur. Unlike other high-reliability industries such as aviation or nuclear power generation, in diving no large literature of diving-specific safety research provides information on human error. Furthermore, the diving literature that does exist tends to focus on equipment design, the effects of diving on cognitive and psychomotor performance, or recreational diving (see Nevo and Breitstein, 1999, for a summary).

The types of behavior necessary for safe and effective performance in a technical context, but are not directly related to technical expertise or psychomotor skills (e.g. decision making, leadership, etc.) are not explicitly addressed as part of U.S. Navy diver training. The *U.S. Navy Diving Manual* (Commander, Naval Sea Systems Command, 1999), the source of Navy diver training and guidance for carrying out diving operations, provides little reference to the specific human factors pertinent to diving. However, U.S. Navy divers may receive instruction in human factors skills, such as leadership or decision making, as part of general military training.

DIVING MISHAP REPORTING

Whenever a diving mishap occurs in the U.S. Navy, a diving mishap report must be sent to the Navy Safety Center (Chief of Naval Operations, 2001). Mishap reports are completed by an individual (usually the Master Diver, Diving Officer, or Diving Medical Officer) involved in treating the injured diver at the command where the accident occurred. The report requires information about the types of injuries, the underwater breathing apparatus (UBA) used, the purpose of the dive, and it gives a narrative description of the accident. The report also includes a system for classifying the cause of the accident.

The diving mishap classification system consists of four categories: procedural factors (i.e. the procedures used were too complex, incorrect, not available), material factors (i.e. the equipment failed due to improper repair, use, or normal wear and tear), design factors (i.e. some aspect of equipment design caused the accidents), and human factors. The individual responsible for completing the mishap report is given the opportunity to identify "*as many causes* [of the mishap] *in each category you determine to apply*" (Chief of Naval Operations, 2001: A6-M-4) and to briefly narrate the events pertaining to the mishap.

If a diving mishap is classified as caused by human factors, a hierarchical structure of four categories and nine subcategories based on the human factors analysis and classification system (HFACS) is used (see Figure 2 for a list of categories and subcategories). HFACS was developed from more than 300 naval aviation incidents obtained from the U.S. Naval Safety Center and has since been refined with data from other military (the U.S. Army Safety and the U.S. Air Force Safety Centers) and civilian organizations (the National Transportation Safety Board, the Federal Aviation Administration; Wiegmann & Shappell, 2001).

Sample

The U.S. Navy Safety Center sent a total of 263 dive mishap reports of accidents that had occurred during operational dives from 1993 to 2002 for examination by the Navy Experimental Diving Unit.

Type of diving mishap

Figure 1 summarizes the type of diving mishap recorded in the report. Decompression sickness (DCS, commonly known as "the bends" and caused by gas bubbles from insufficient decompression) was the most common diving injury reported (122 cases), and arterial gas embolism (AGE; 97 cases, caused by gas bubbles entering into the arterial circulation and then acting as blood vessel obstructions called emboli) was the second most frequent mishap. "Oxygen toxicity" (5 cases) covered any body injury that resulted from inhaling too much oxygen: above atmospheric pressures, oxygen can affect the central nervous system and cause seizures and convulsions. Five cases of near drowning were included: these can occur through overexertion, panic, or exhaustion and can be considered mishaps that had almost resulted in death.

Five deaths had occurred in Navy diving during the reporting period, and these are discussed in detail in the section concerned with diving fatality reports. The 29 "other" injuries included: 10 mechanical injuries, nine cases of non-AGE pulmonary overinflation syndrome (caused by gas expanding within the lung), eight incidents of asymptomatic missed decompression (in which the diver, despite not having any symptoms, was treated for DCS, because a certain amount of decompression time was known to have been omitted), and two instances of chemical burns (resulting from a caustic solution introduced into the upper airway of a diver from a UBA's carbon dioxide scrubber).

Figure 1. Mean incidents of 263 U.S. Navy diving mishaps from 1993 to 2002 (whiskers represent a 95% confidence interval).

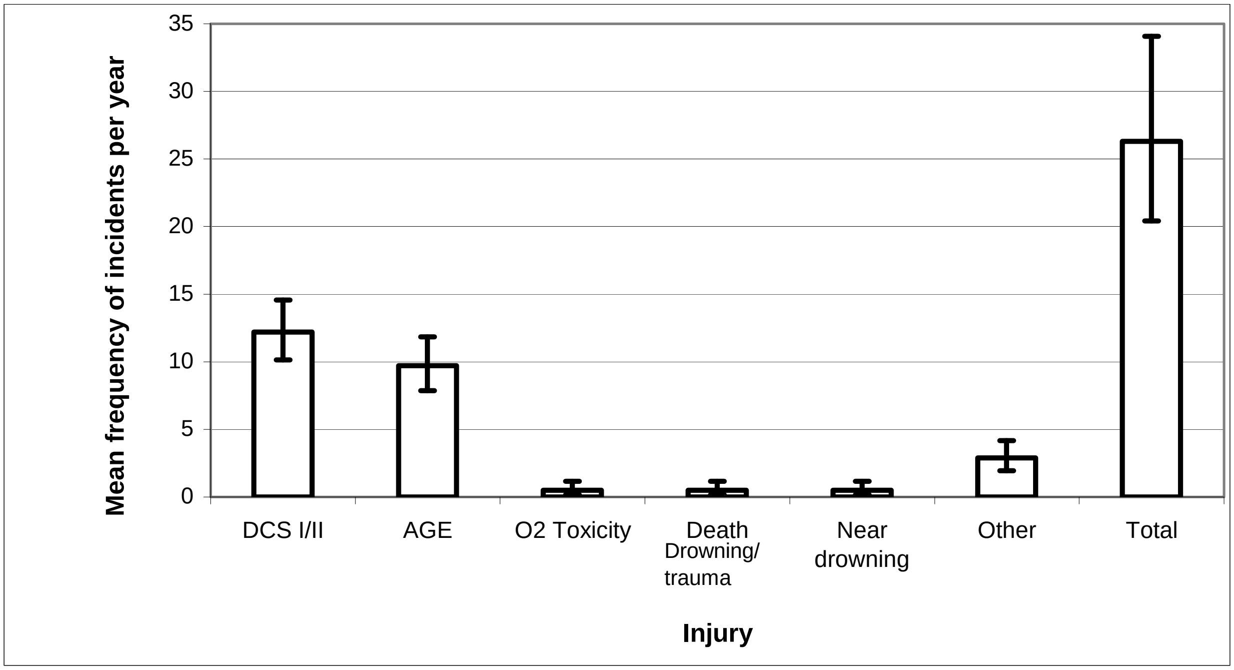


Figure 1 presents the mean frequency of each type of incident per year. The whisker plots in Figure 1 represent a 95% confidence interval based upon a Poisson distribution. The Poisson distribution provides a good model of data in which the probability that an event (i.e., an accident) will occur is low, all events are independent, and the average rate does not change over the period of interest. The Poisson distribution is frequently used to model accident data (Nicholson & Wong, 1993).

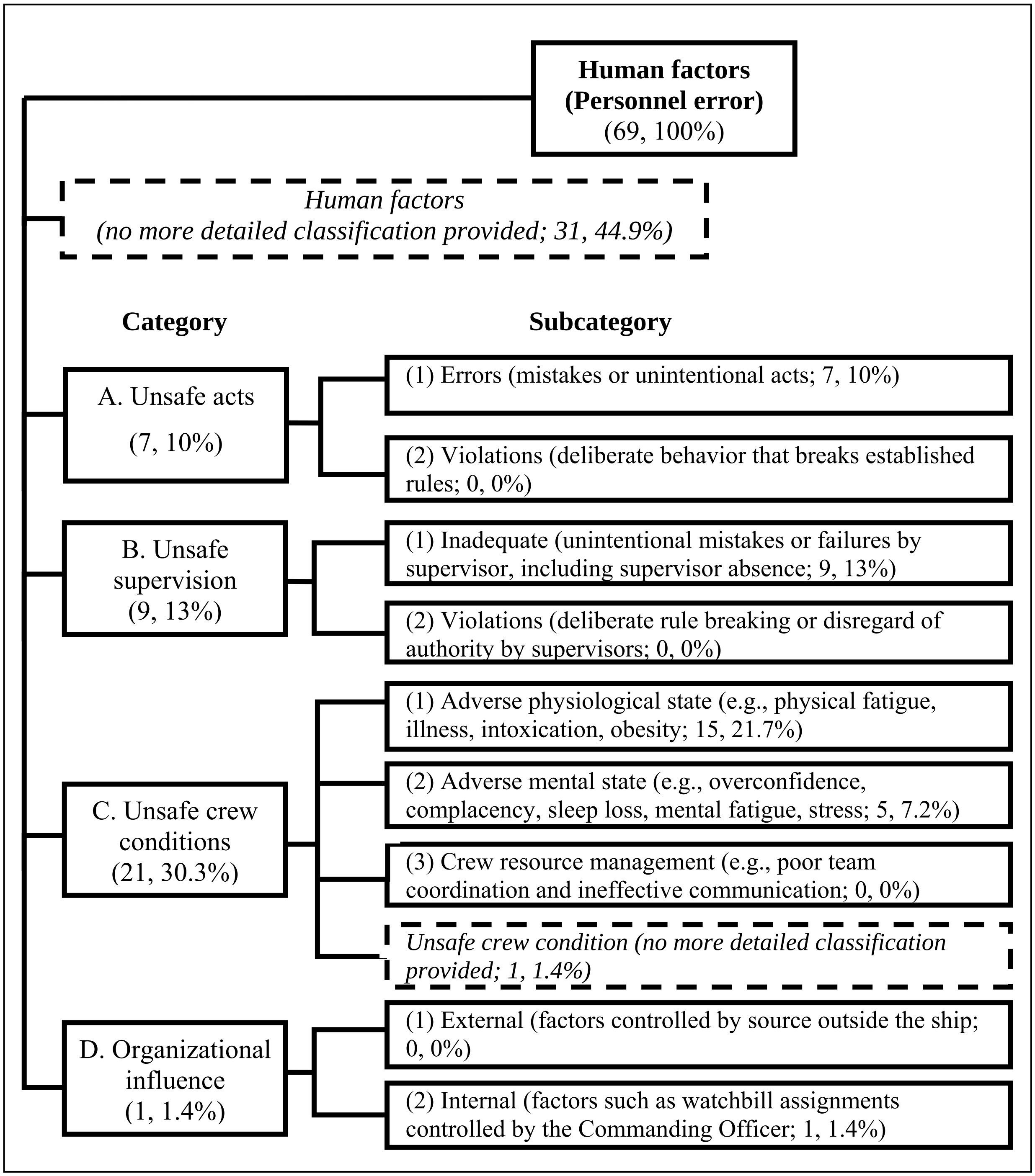
Classification of the causes of diving mishaps

From the 263 reports, a total of 185 diving mishaps (70%) were attributed to “unknown” causes by the persons responsible for completing the mishap reports. It should be noted that “unknown” is not included as a possible category in the diving mishap classification system, nevertheless “unknown” was still recorded as the cause on the mishap report. None of the categories from the diving mishap classification system were used to categorize the causes of those mishaps. Of the 78 remaining dive mishaps, 60 (23%) were attributed to human factors, eight (3%) to material factors, five (2%) to procedural factors, and five (2%) to environmental factors. Procedural factors are defined as the possible effect of regulations, operations, and processes. However, the guidance to accident investigators explicitly states that a person not following the procedures should be recorded as a human factor, and not a procedural factor (Chief of Naval Operations, 2001).

Human factors classification of diving mishaps

For the 60 dive mishaps that were attributed to human factors— out of the 263 reports examined—69 accident causal factors were identified. The Chief of Naval Operations (2001) outlines the structure for coding the human factors causes of a diving mishap. The system is a hierarchical structure of four categories and nine subcategories. Figure 2 depicts how the 69 causal factors attributed to human error were categorized at a greater level of detail.

Figure 2. Using the dive mishap human factors classification system to categorize the 69 human factors for accident causation identified from 60 diving mishaps.



From Figure 2 it can be seen that of these 69 causal factors, only 38 (55.1%) were further classified into the four HFACS categories and underlying subcategories. A total of 31 (44.9%) were simply labeled as human factors, with none of the more detailed subcategories used to classify the mishap. To illustrate, in one diving mishap, a case of DCS was categorized as "human factors." However, reading the narrative indicates that the diver was fatigued and dehydrated, both of which are conditions regarded as risk factors for DCS. It seems that this diving mishap could have been more specifically classified as due to human factors/unsafe crew conditions/adverse physiological state causes.

To summarize, despite the instruction stating that accident "*causes should be one of the four major categories*" (Chief of Naval Operations, 2001: A6-M), with 'unknown' not being provided as one of the categories the majority of dive mishap causes were classified as unknown (an explanation for the use of 'unknown' is presented in the discussion). Moreover, even in cases for which causes are known, they are most commonly attributed to unspecified human factors. When the type of human error is specified, it is most commonly attributed to inadequate supervision and an adverse physiological state (e.g., physical fatigue, illness, intoxication). Possible reasons for the failure of the mishap investigators to use the mishap categorization system are outlined in the discussion. Therefore, to realize the objective of identifying human errors in U.S. Navy diving accidents, it was necessary to identify a rich source of accident data and develop a system to categorize the active human factors failures.

DIVING FATALITY REPORTS

Five Mishap Investigation Reports (MIRs) from fatal U.S. Navy diving accidents were identified for analysis. These five accidents were not chosen by the researchers; the U.S. Navy Supervisor of Diving selected these as the last five U.S. Navy diving fatalities. MIRs are required to answer the who, what, where, when, and why questions about on-duty diving accidents that result in a Navy diving fatality (Chief of Naval Operations, 2001). Developed from interviews with the individuals involved, from attempts to save stricken divers, analyses of the equipment, and dive mishap reports, the MIRs contain detailed accounts of these incidents. Of the five fatalities, three were attributed to drowning, one to AGE, and another to trauma.

Since the focus of the study was to identify active failures that are made by members of the dive team and that result in an accident, many current classification systems were examined to assess their utility in helping to achieve this goal. The criteria for selecting a classification system were based upon whether it allowed a detailed classification of active failures, applicable to the operations performed by a dive team, and whether the categories were sufficiently distinct to allow acceptable levels of interrater reliability. Reporting systems based on Reason's 1990 model of accident causation include HFACS (Wiegeman & Shappell, 2001), Tripod (Groeneweg, Lancioni, Metaal, 2003) and the incident cause analysis method (ICAM; Hayward, Lowe, & Gibb, 2002). However, these classification systems do not provide a detailed system to categorize the human errors made by the members of the dive team. To illustrate, HFACS has a crew resource management subcategory, but the specific active failures such as situation awareness or communication are not delineated.

One system that did seem to provide sufficient detail to account for the active failures committed within diving was the Offshore Nontechnical Skills Framework (O'Connor and Flin, 2003). Although this framework is not an accident classification system, it outlines the individual skills required by effective offshore production personnel. The framework consists of six categories: situation awareness, decision making, communication, teamworking, personal resource, and supervision/leadership. Each of these categories is further divided into between three and five specific skills. This O'Connor and Flin framework was based on research into human factors carried out in the offshore oil industry (e.g., Flin, Mearns, Gordon, & Fleming, 1996; O'Dea & Flin, 2001; Parkes, 1992; Rundmo, 1994) and the NOTECHS behavioral marker system.

Behavioral markers are "*a prescribed set of behaviors indicative of some aspect of performance*" (Flin & Martin, 2001: 96). The NOTECHS behavioral markers system consists of four independent categories of behaviors (cooperation, leadership and management skills, situation awareness, and decision making). Each category is then further subdivided into three or four elements. These four categories of behavior were chosen on the basis of theoretical models identified from a literature review (see Avermaete & Kruijsen, 1998) and a review of existing behavioral marker systems (Flin & Martin, 2001).

However, the O'Connor and Flin framework was based on research carried out in the offshore oil production and the aviation industries, rather than on diving. Furthermore, the framework was not designed specifically to classify human error but to provide a structure to delineate the individual skills required by personnel for safe and productive operations. Thus, it was decided not simply to classify the MIRs with the O'Connor and Flin framework, but to use it as a guide for developing an error taxonomy based on the diving data.

Consensual Qualitative Methodology

A method based on the CQR technique (Hill, Thompson, & Williams, 1997) was employed to guide the adaptation of the O'Connor and Flin framework into a dive team human error taxonomy using the data from the MIRs,

Hill et al (2005) states that CQR is ideal for studying events that are hidden from public view, that are infrequent, that occur at varying time periods, that have not been studied previously, or that have had no measures created for them.

CQR is a qualitative methodology that incorporates elements from

- grounded theory (the theory is devised from data systematically gathered through the research process; Strauss & Corbin, 1998),
- phenomenology (exponents of this approach are interested in the qualitatively different ways in which phenomenon occur; Giorgi, 1985), and
- comprehensive process analysis (this is a narrative research method for describing and understanding the significance, effects, and context of clinical events; Elliot, 1989).

Three general steps described by Hill et al (1997) for carrying out CQR were used to develop a human error taxonomy for diving:

1. Domains. The first stage of the process was to divide the account of the fatality mishap into domains (i.e., topic areas). A “start list” (Miles & Huberman, 1994) of domains derived from the O’Connor and Flin nontechnical skills framework for offshore operations was used as an initial structure.

Two raters (one of whom is a U.S. Navy diver) — both with Ph.D.’s in industrial/organizational psychology and with experience investigating accidents in the offshore oil and nuclear power generation industries — carried out the data analysis. Each rater read through the five diving fatality reports and independently developed a set of domains. These domains were compared, and areas of agreement and disagreement were discussed. The six overarching categories of the O’Connor and Flin framework were found to be comprehensive enough to account for the diving data. However, in order to accommodate the diving data, it was necessary to adapt the framework at the subcategory level by developing a set of discrete subcategories that would result in high levels of interrater agreement. The alterations made to the O’Connor and Flin framework were:

- Four subcategories- ‘procedural adherence’, ‘experience/training’, ‘choice of leadership style’, and ‘appropriate use of authority’ - were added as there was not an existing subcategory to describe the data.

- ‘Giving appropriate feedback’, ‘asking questions’ ‘listening’, and ‘attending to nonverbal signals’ were merged to form a new subcategory called ‘information exchange’.
- ‘Environmental awareness’ was merged as part of ‘risk and time assessment.’
- ‘Considering other’, ‘supporting others’, and ‘team decision making’ were combined as an additional subcategory entitled ‘team climate.’
- ‘Shared mental models’ was renamed and expanded as part of the new ‘task awareness’ subcategory.
- ‘Maintaining team focus’ was included as part of the ‘concentration/avoiding distractions’ subcategory.

2. Core ideas. The data was “edited” (as Hill et al [1997] have suggested) into a standard format that was concise, clear, and comparable across the five MIRs. The events were put into chronological order, repetitions omitted, and a single account of the accident was written.

3. Cross analysis. A total of 152 statements were identified for categorization from the five MIRs, and the raters independently categorized each of the statements. If a statement could be classified into more than one subcategory, then the raters categorized the statement into what they deemed to be the primary category. This parsimony of coding was necessary to obtain acceptable levels of reliability between the raters.

When the coding was first completed, the interrater reliability was Cohen's $\kappa = 0.60$, and the two raters examined the 54 statements for which there was a disagreement. The disagreements were generally attributed to those statements that could possibly be classified in more than one category. These judgments were discussed, the raters then independently recoded the statements, and a Cohen's κ = 0.81 resulted. Landis and Koch (1977) propose that values over 0.81 indicate near perfect agreement. The raters discussed those 31 statements to which they did not agree, until they reached a consensus about which subcategory in which to place them.

Table 1 shows the frequency and percentage for each category and subcategory used to classify the human errors, a frequency and percentage that represent the sum of the classifications made at the sub category level. It also provides a characteristic example of a statement that was coded into each of the subcategories. The numbers in parentheses in Table 1 represent the percentage of good examples of the behavior where the dive team worked well to recover from an incident. For each category,

- *Situation awareness* refers to the behaviors by which individual team members build and share mental pictures of the situations to create a common understanding.
- *Decision making* is concerned with following, and using, the procedures for carrying out a task and reviewing the outcomes of a solution to assess whether the goal has been reached.
- *Communication* concerns the sharing of information among team members.
- *Team cohesion* concerns behaviors indicating a sense of "teamness" among the team members.
- *Personal resources* refers to any factor that reduces an individual's level of performance: e.g., stress, fatigue, physical or mental fitness, and lack of experience or training.
- *Supervision/leadership* includes the direction and structure provided by both the leader and other team members.

Table 1. Dive team human error taxonomy.

Category	Skill	Description	Example obtained from interviews	Fatalities (n=5) Freq	Fatalities (n=5) %	Interviews (n=15) Freq	Interviews (n=15) %
Situation Awareness	Anticipation	Forward planning is completed to identify and discuss contingency strategies and/or possible future problems.	*Started screw change, he had done a dozen in the past.*	10 (1)	6.5 (0.7)	2 (2)	1.1 (1.1)
	Problem definition/ diagnosis	Information is gathered to identify a problem and its causal factors.	*They did not want to admit something had gone bad.*	1	0.7	8	4.4
	Risk and time assessment	An accurate assessment of risk and time is completed (weather, sea state, time available, equipment, etc.)	*There was some concerns about the moorings, but they decided to continue anyway.*	30 (5)	20.1 (3.2)	35 (1)	19.1 (0.6)
	Dive status awareness	Every team members has an accurate awareness of how a dive is progressing.	*After about 30 minutes, getting no line-pull signals.*	12 (2)	7.8 (1.4)	21	11.5
	Task awareness	The team member has an accurate awareness of the task in which he/she is engaged and of his/her role in the dive.	*Told he was red diver, then told he was yellow. Went over to yellow bike, then told he was red diver.*	7	4.5	6 (2)	3.4 (1.1)
	Concentration/ avoiding distraction	The team member is able to give the attention necessary to perform the task.	*The supervisor said to load a HeO_2 bottle, but he loaded an N_2O_2 bottle instead.*	0	0	4 (1)	2.3 (0.6)
	Total			**60 (8)**	**39.6 (5.3)**	**76 (6)**	**41.8 (3.4)**
Decision Making	Procedural adherence	The procedures are followed correctly and are appropriate for the task being carried out.	*The correct procedure was not followed to check the equipment.*	5	2.6 (0.7)	5	2.8
	Outcome review	The outcome of a solution is checked against the predefined goal.	*Nothing was learned from the incident.*	0	0	2	1.1
	Total			**5**	**2.6 (0.7)**	**7**	**3.9**
Communication	Assertiveness/ speaking up	Ideas and observations are communicated in a manner that is persuasive to other team members.	*He suggested diving would not be a good idea but did not want to push it, as he knew he would be overruled.*	4	2.6	5 (1)	2.8 (0.6)
	Information exchange	Information is clearly and accurately exchanged between team members.	*He would not listen to recommendations.*	5 (1)	3.2 (0.7)	10	5.7
	Total			**9 (1)**	**5.8 (0.7)**	**15 (1)**	**8.5 (0.6)**

Table 1. (continued)

Category	Skill	Description	Example obtained from interviews	Fatalities Freq	Fatalities %	Interviews Freq	Interviews %
Team cohesion	Team climate	Team members are aware of the competencies of their teammates, trust each other, and have a positive attitude toward being a member of the team.	*He requested extra divers, but they were reluctant to work outside normal working hours.*	2 (1)	2 (0.7)	8	4.6
	Conflict solving	Conflicts are resolved in a way that minimizes harm done to both parties.	*The shipyard viewed them as taking their work and so were not particularly helpful.*	0 (1)	(0.7)	1	0.6
	Total			**2 (2)**	**2.0 (1.4)**	**9**	**5.2**
Personal resources	Identifying & managing stress	Signs of stress are communicated and taken into account.	*There was a lot of tension topside.*	6	3.9	6 (1)	3.4 (0.6)
	Identifying & managing fatigue	Signs of fatigue are communicated and taken into account.	*They were diving around the clock, so the divers were tired.*	4 (1)	2.6 (0.7)	4	2.3
	Physical & mental fitness	Team members are sufficiently fit, physically and mentally, to perform the assigned tasks.	*He felt as if he was in a daze.*	1	0.7	2	1.1
	Experience/ training	The team members involved in the operation have sufficient experience and training.	*For many of the divers, it was the first time in a dry suit.*	12	7.8	12 (2)	6.8 (1.1)
	Total			**23 (1)**	**15.0 (0.7)**	**24 (3)**	**13.6(1.7)**

Table 1. (continued)

Category	Skill	Description	Example obtained from interviews	Fatalities Freq	Fatalities %	Interviews Freq	Interviews %
Supervision/ Leadership	Appropriate use of authority	The supervisor adequately balances assertiveness and team member participation.	*He told the divers they were a bunch of pussies and he would get in the water if they were too scared.*	3	1.9	5	2.8
	Maintaining standards	The supervisor ensures the dive team complies with standard operating procedures and intervenes if required.	*They should have used the MK21* [a particular type of dive helmet] *for this evolution.*	16	10.4	19	10.2
	Planning and coordination	The appropriate personnel, resources, and techniques are selected to complete a task.	*There was some question about splashing a new diver who had never been in these conditions before.*	18	12.3	6	3.4
	Workload management	Tasks and resources are shared in order to achieve top performance and avoid workload peaks and dips.	*They knew there were jobs piling up behind them.*	2	1.3	4	2.3
	Choice of leadership style	A leadership style is used that promotes a safe working environment and is appropriate to the dive team, task, and urgency of the situation.	*The MDV was a cowboy.*	2	1.3	5	2.8
	Total			**41**	**27.2**	**39**	**21.5**
	Grand Total			140(12)*	91.5(8.5)	170 (10)	93.5 (6.5)

*Numbers in parentheses indicate good examples of the behavior.

Poor leadership, poor situation assessment (particularly risk and time assessment), poor supervision/leadership, and lack of personal resources were the most commonly used categories and subcategories.

The fatality MIRs were found to be a rich source of data for developing the dive team human error taxonomy and for identifying the active failures that result in accidents. However, fatal accidents represent only a small subset of the most extreme accidents that occur in diving. Research in industries other than diving suggests that as many as 600 near misses may occur for every 10 minor injuries and for every serious injury (Bird & Germain, 1996). Therefore, expanding the data set of accidents investigated to include nonfatal accidents and near misses was needed.

CRITICAL INCIDENT INTERVIEWS

Originally developed to examine flight crew selection, readiness, and performance (Flanagan, 1954), CIT interviews were employed to elicit accounts from U.S. Navy divers who had either been involved in a diving accident as a victim or a member of the dive team or had been involved in a situation that they felt could have resulted in an accident (i.e., a near miss). By using a respondent's recollection of a specific incident as its starting point and employing a semistructured interview format involving several "sweeps" through the incident. The four sweeps in the CIT of divers included:

Sweep 1 — The interviewees were prompted to identify a diving accident or near miss in which they had been involved and articulate it. Each participant was asked to describe the event from his/her own perspective — to describe it in detail, stage by stage, as it had developed.

Sweep 2 — Filling in gaps in the incident. The interviewer repeated the situation back to the respondent in order to check the interviewer's understanding. This sweep helped to pinpoint gaps both in time and events and typically aided respondents in recalling the missing portions.

Sweep 3 — Expanding on the incident. The interviewer reviewed the event again, this time probing at various points and asking for more detailed descriptions of the human errors that occurred during the situation. This sweep involved questioning the reasoning and looking for cues and rationales for the actions taken by members of the dive team.

Sweep 4 — Posing "what-if" queries. A typical question was, "Would you have acted the same way at an earlier point?" or "Would someone with less experience than you have acted in a similar way?" The researcher listened for other possible courses of action and interpretations.

A total of 15 CIT interviews were completed with U.S. Navy divers. Two of the incidents had resulted in a death, one in a near drowning, another in a serious physical injury, and the remaining 11 were near misses in which no injury resulted, although the potential for death or serious injury had been high. Rather than transcribing the interviews, the researchers used a digital recording and interview notes to develop a single detailed account of the incident.

Two of the accidents discussed in the interviews had occurred between 1993 and 2002, so it was possible to examine the relevant dive mishap reports. Neither of these reports disagreed with the accounts from the interview. However, they also did not provide any additional information than that obtained from the interview, despite the fact that the reports were drawn from information collected from a number of individuals.

Data analysis

The framework of six categories and 21 subcategories developed from the MIRs and outlined in Table 1 were used to classify the statements from the 15 CITs. Therefore, it was not necessary to complete the first stage of the CQR (i.e., to develop the domains). However, the next two stages of the CQR were completed in the same manner as for developing the diving human error taxonomy. A total of 180 statements were identified from the CIT interviews and categorized by the same two raters used to develop the original coding system.

When the taxonomy was first used to classify the statements from the CIT interview, the interrater reliability was Cohen's $\kappa = 0.60$. After a review of areas of agreement and disagreement between the coders, the statements were recoded and an interrater reliability of Cohen's $\kappa = 0.86$ was achieved. The 20 statements on which there was no agreement were discussed and placed into an agreed upon subcategory. Table 1 shows that the proportion of statements assigned to each category and subcategory was broadly similar to that for the five MIRs. The taxonomy was found to account for all the factors identified through the CIT interviews.

DISCUSSION

U.S. Navy diving mishap reporting

Analysis of the dive mishap reports reveals that the largest proportion of the mishaps (70%) are attributed to unknown causes, and only 23% are attributed to human factors. The proportion of causes attributed to human error is far below the 80% generally attributed to human error in high-reliability industries (Reason, 1990). Four possible reasons may account for this finding.

First, a certain proportion of mishaps in diving, unlike those in any other industry, are expected. Even if decompression tables are used correctly, a small number of DCS cases are expected to occur. To illustrate, the mean risk of receiving a case of DCS when U.S. Navy no-decompression air tables are used is 2.1% (Gerth & Thalmann, 2000). However, in real-world operations this DCS rate is not realized, as diving supervisors will generally not dive at the "edges" of the dive tables. For example, the no-decompression limit for a 60-foot air dive is 60 minutes (Chief of Naval Operations, 2001). Unless there is an operational necessity, however, a diving supervisor will normally instruct the divers to abort the dive after a maximum of 50 minutes. As a result the risk of DCS to the divers is reduced.

A second possible reason for a lack of human factors causes in diving accident reports is a lack of understanding about what “human factors” actually denotes. No formal training is provided in classifying diving mishaps, and the individual completing the investigation is likely to have little prior experience of investigating such mishaps.

A third possibility is a reluctance to provide a detailed, accurate account of mishaps: narratives of mishaps can vary from a few sentences to a couple of paragraphs, and this reluctance is evident in the great variation in the quality of the reports. Generally there is little discussion of the mishap itself; rather, the reports tend to outline the treatments of the injuries.

Another possible reason that diving mishap reports may lack human factors causes is that the HFACS may not be not applicable for this purpose. Just as there have been problems with taking CRM training developed for one domain and applying it to another (Flin, O’Connor, & Mearns, 2002; Helmreich, 2000), the same may be true of applying a human error taxonomy to a domain for which it was not developed.

The collection of accurate accident data is important for improving industrial safety. However, research suggests that many accident reporting systems are vulnerable to the kinds of difficulties described. Underreporting, incomplete recordings, and incomplete information about conditions and contexts are common to many systems and do not provide a complete picture of the conditions under which accidents result (Stoop, 1997). Gordon, Flin, and Mearns (2005) state that most oil companies who operate on the U.K. continental shelf may produce accident forms with extensive information, but the quantity and quality of the data concerning the human factors causes of such accidents is generally poor.

Only experienced individuals, knowledgeable about both the domain in which the accident has occurred and human factors, should delineate the causes of accidents. The findings from the analysis of diving mishap reports illustrate that the users of recording systems need to be trained to classify human error accurately. The training should be designed to ensure that investigators are able to use the taxonomy accurately and consistently, by reducing the likelihood of judgment biases and improving inter-rater reliability. A large body of research (e.g., Baker, Mulqueen, & Dismukes, 2001) exists on training observers to reliably use behavioral markers. It may be possible to adapt some of the principles of such training to instruct accident investigators in using human error taxonomies.

Human error in U.S. Navy diving

The pattern of human errors identified from the fatality reports was broadly similar to those identified from the CIT interviews. Situation awareness (particularly risk and time assessments) was the category most commonly used for classifying the human error (see Table 1). It is necessary for each member of a team not only to understand what he or she is doing but also to know what other members of the team are, or are not, doing (Endsley & Robertson, 1996).

That leadership failures were also evident (see Table 1) is not surprising. Supervisors hold huge influence on issues such as compliance with safety rules (Thompson, Hilton, & Witt, 1998). North Sea divers have reported that their confidence in the supervisor's ability to manage accident risk was the most important factor in preventing accidents (Osman, Adie, & Cairns, 2003). Thus, the supervisor's ability to manage risk and the attitudes and leadership skills of those in the supervisory position are crucial to the safety and effectiveness of a diving mission.

U.S. Navy divers also recognize the significance of human error in diving operations. When asked to identify the three main causes of U.S Navy diving accidents, 272 recently surveyed U.S. Navy divers identified human failure rather than mechanical or environmental causes as the main sources of diving accidents (O'Connor, 2005). The five most common causes they identified were (1) complacency, (2) fatigue, (3) inexperience, (4) training, and (5) planning.

CONCLUSION

This study has shown that researchers should be cautioned about using human error taxonomies for identifying and classifying human error in one industry and applying them to another. Although there will be similarities at the category level with common factors such as situation awareness, decision making, communication, and leadership, the same is not likely to be as true at the subcategory level. There may be a necessity for subcategories that are specific to the industry (e.g. dive status awareness), or a requirement for more or less subcategories dependent upon the types of accidents that are occurring. On the basis of developing the dive team human error taxonomy, it is possible to propose a number of considerations that should be addressed when developing, or adapting, an error taxonomy from one industry to another.

The difficulty of studying accidents in most high-reliability industries is that they happen infrequently. The information available to researchers on the small number of accidents that do occur may be limited, inaccurate, or focused on technical issues. The study described in this paper indicates that researchers must recognize the errors and inconsistencies in the data contained in accident reports. Researchers must examine ways in which the accident report is completed, consider the culture of the organization (e.g., is there a reporting culture?), and consider the level of accident investigation training among those completing the reports.

Researchers should be especially cautious when applying an error taxonomy developed for use in an industry with a high level of safety culture maturity (e.g. aviation, nuclear power generation) and applying it to an industry with a less developed safety culture (e.g. fishing, construction). Additionally, the same is true when using an error taxonomy developed in one industry (e.g. railway) and applying it to a very disparate industry (e.g. offshore oil production).

Researchers should be wary of industry specific language that is used in error taxonomies to identify a particular category or subcategory. To illustrate, crew resource management is a widely used and understood concept in the aviation industry. However, the converse is true in Navy diving, so it is unsurprising that the crew resource management subcategory was not used to identify any of the causes of the 263 diving accidents examined.

The researcher must strike a balance between an error taxonomy that is so small that it provides insufficient information, and so big that it is unwieldy and has low levels of reliability. Researchers should endeavor to construct a taxonomy that is parsimonious, and consists of discrete subcategories that are sufficiently distinct to lead to high levels of interrater agreement.

To improve reliability between accident investigators, and ensure that as much information as possible is collected about an accident, it is suggested that investigators are encouraged to utilize a bottom-up approach to error classification by starting the classification at the subcategory level. This means that the accident investigator does not need a thorough understanding of category level concepts such as decision making and situation awareness to use the error taxonomy appropriately. However, for the bottom-up approach to be successful, the classification system must be sufficiently parsimonious, the subcategories adequately distinct, and the error taxonomy properly tailored to the specific industry.

The combination of the CIT and CQR methods is a technique for researchers to develop or customize error taxonomies for those industries for which no large body of human error research provides guidance. The CIT offers the researcher a method for obtaining a rich source of data regarding accidents or near misses. A small sample size can provide insight into the causes of the accident or near miss, as well as into the culture of the organization.

Certain limitations to the CIT are recognized. The people interviewed will describe only what can be verbalized, a consequence that results in overlooking information which cannot be verbalized or is difficult to verbalize. A further problem is that the respondents may inaccurately justify their decisions when describing their actions after the event. Finally, this CIT is very labor intensive: it may take at least three times as long to sort and assimilate the data as it does to obtain it (Sinclair, 1992). Nevertheless, the CIT technique is a method for collecting qualitative data with sufficient structure to provide some degree of consistency between different interviews and for giving the researcher the freedom to obtain more information about particular aspects of a specific accident or near miss.

Qualitative research "*not only generated large amounts of data, but it generates data in a non-standard format. . . . It is not uncommon for researchers to find themselves overwhelmed by such a large amount of data*" (Turner, 1983: 333). The CQR method gives a technique for organizing qualitative data collected from accident reports or critical incident interviews with individuals who have been involved in an accident or have had a near miss. The CQR provides a structured system for organizing a large amount of data, allows the researcher to amalgamate and organize the data so that conclusions can be drawn from the information, and provides a method for obtaining acceptable levels of interrater reliability. However, as with the CIT, the CQR method is extremely time intensive and requires a minimum of two researchers knowledgeable about human factors and the industry in which they are working.

The need to accurately identify the frequency and types of human errors that can cause accidents is necessary for any organization striving to have an advanced safety culture. We suggest that CIT interviews and the CQR technique provide safety researchers with a rigorous methodology for carrying out qualitative research to meet this requirement.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Denis Thomas for his editorial comments. The information presented is the opinion of the authors and does not necessarily reflect that of the U.S. Navy.

BIBLIOGRAPHY

Avermaete, J.A.G., & Kruijsen, E. (Eds.). (1998). *The evaluation of non-technical skills of multi-pilot aircrew in relation to the JAR-FCL requirements. (Project report: CR-98443)*. Amsterdam, The Netherlands: NLR: EC NOTECHS.

Baker, D., Mulqueen, C., & Dismukes, R.K. (2001). Training raters to assess resource management skills. In E. Salas, C. A. Bowers, & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 131–146). Mahwah, NJ: Lawrence Erlbaum Associates.

Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error* (pp. 271–283). Chichester: John Wiley.

Bird, F.E. & Germain, G.L. (1996). *Practical loss control leadership*. Loganville, GA: Det Norske Verita.

Chief of Naval Operations (2001). *Navy occupational safety and health (NAVOSH): Program nanual for forces afloat OPNAVIST 5100.19D*. Washington, DC: Department of the Navy.

Commander, Naval Sea Systems Command (1999). *U.S. Navy Diving Manual, Revision 4*. Arlington, VA: Naval Sea Systems Command.

Elliot, R. (1989). Comprehensive process analysis: Understanding the change process in significant therapy events. In M.J. Packer and R.B. Addison (Eds.), *Entering the circle: Hermaneutic investigations in psychology* (pp. 165–184). Albany, New York: SUNY Press.

Endsley, M. R. & Robertson, M.M. (1996). Team situation awareness in aircraft maintenance. In Federal Aviation Administration (Ed.), *Phase VI Progress Report*. Washington, DC: FAA.

Flanagan, J.C. (1954). The Critical Incident Technique. *Psychological Bulletin*, 51(4), 327–358.

Flin, R., & Martin, L. (2001). Behavioral markers for CRM: A review of current practice. *International Journal of Aviation Psychology,* 11, 95–118.

Flin, R., Mearns, K., Gordon, R. & Fleming, M. (1996). Risk perception by offshore workers on UK oil and gas platforms. *Safety Science,* 22, 131–145.

Flin, R., O'Connor, P., & Mearns, K. (2002). Crew resource management: Improving safety in high reliability industries. *Team Performance Management*, 8, 68–78.

Gerth, W.A., & Thalmann, E.D. (2000). Estimated DCS risks of “reverse” dive profiles. In M.A. Lang and C.E. Lehner (Eds.) *Proceedings of reverse dive profiles workshop*, (pp. 145–170). Washington, DC: Smithsonian Institution.

Giorgi, A. (1985). Sketch of a psychological phenomenological method. In A. Giorgi (Ed.), *Phenomenology and psychological research* (pp. 8–22). Pittsburgh, PA: Duquesne University Press.

Gordon, R., Flin, R. & Mearns, K. (2005). Designing and evaluating a human factors investigation tool (HFIT) to improve incident investigation. *Safety Science*, 43, 147–171.

Groeneweg, J., Lancioni, G.E., Metaal, N. (2003). Tripod: managing organizational components of business upset. In T. Bedford & P.H.A.J.M. van Gelder (Eds.) *Safety and Reliability: Proceedings of the Esrel 2003 Conference* (pp. 707-712), Holland: Aa Balkema,

Hayward, B., Lowe, A., & Gibb, G. (2002). ICAM: Applying reason to safety investigation. Paper presented at the 25th Annual Conference of the European Association for Aviation Psychology: Safety, Systems and People in Aviation, Warsaw, Poland.

Helmreich, R.L. (2000). On error management: Lessons from aviation. *British Medical Journal*, 320, 781–785.

Hill, C.E., Knox, S., Thompson, B.J., Williams, E.N., Hess, S.A., & Landany, N. (2005). Consensual qualitative research: An update. *Journal of counseling psychology*, 52(2), 196–205.

Hill, C.E., Thompson, B.J., & Williams, E.N. (1997). A guide to conducting consensual qualitative research. *The Counseling Psychologist*, 25, 207–217.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.

Miles, M.B. & Huberman, A.M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.

Nevo, B. & Breitstein, S. (1999) *Psychological and behavioral aspects of diving*. Flagstaff, AZ: Best Publishing Company.

Nicholson, A., & Wong, Y.D. (1993). Are accidents Poisson distributed? A statistical test. *Accident Analysis and Prevention*, 26(3), 407–408.

Norman, D.A. (1988). *The design of everyday things*. New York: Doubleday Currency.

O'Connor, P. & Flin, R. (2003). Crew resource management for offshore oil production teams. *Safety Science*, 41, 591–609.

O'Connor, P (2005). *An investigation of the nontechnical skills required to maximize the safety and productivity of U.S. Navy divers*. Panama City, FL: Navy Experimental Diving Unit.

O'Dea, A., & Flin, R. (2001). Site managers and safety leadership in the offshore oil and gas industry. *Safety Science,* 37, 39–57.

Osman, L.M., Adie, W. & Cairns, J. (2003). *Attitudes to safety culture among professional divers and offshore workers.* Suffolk, U.K.: HSE Books.

Parkes, K.P. (1992). Mental health in the oil industry: A comparative study of onshore and offshore employees. *Psychological Medicine,* 22, 997–1009.

Reason, J. (1990). *Human error*. Cambridge, U.K.: Cambridge University Press.

Rundmo, T. (1994). Associations between organisational factors and safety and contingency measures on offshore petroleum platforms. *Scandinavian Journal of Work Environment and Health*, 20, 122–127.

Shadbolt, N., & Burton, M. (1992). Knowledge elicitation. In J.R. Wilson & N. E. Corlett (Eds.), *Evaluation of human work: A practical ergonomics methodology* (pp. 321–345). London: Taylor & Francis.

Stoop, J. (1997). Accident scenarios as a tool for safety enhancement strategies in transportation systems. In A. Hale, B. Wilpert & M. Freitag (Eds.), *After the event: from accident to organisational learning* (pp. 77–93). Oxford: Elsevier Science Ltd.

Strauss, A. & Corbin, J.M. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oakes, CA: Sage Publications.

Thompson, R.C., Hilton, T. F., & Witt, L. A. (1998). Where the safety rubber meets the shop floor: A confirmatory model of management influence on workplace safety. *Journal of Safety Research,* 29(1), 15–24.

Turner, B.A. (1978). *Man-made disasters*. London: Wykeham.

Turner, B.A. (1983). The use of grounded theory for the qualitative analysis of organizational behavior. *Journal of Management Studies,* 20(3), 333–348.

Wiegeman, D.A. & Shappell, S.A. (2001). Human error analysis of commercial aviation accidents: Application of the human factors analysis and classification system (HFACS). *Aviation, Space, and Environmental Medicine*, 72(11), 1006–1017.

BIOGRAPHIES

Paul O’Connor is a Navy Aerospace Research Psychologist (Pensacola, Florida). He received his Ph.D. in Industrial/Organizational Psychology from Aberdeen University, Scotland in 2002.

Angela O’Dea is a Senior Scientist at of Applied Research Associates (Klein Associates division, Dayton, Ohio). She received her Ph.D. in Industrial/Organizational Psychology from Aberdeen University, Scotland in 2002.

John Melton is a Navy Research Psychologist (Great Lakes, Illinois). He received his Ph.D. in Counseling Psychology from Loyola University, Chicago in 2002.