

## The Utility of Human Factors Analysis and Classification System (HFACS) in the Analysis of Military Aviation Accidents

Yousuf Al-Wardi; MD, MSc

### Abstract

The Human Factors Analysis and Classification System (HFACS) has proved to be a practicable and reliable method for analysing and classifying human factors in aircraft accidents. The system is based on Reason's (1990) model for active failures and dormant conditions. It explores human factors at operational, supervisory and higher management levels. The aim of this study was to examine the utility and reliability of HFACS system in identifying the human factors contributing to aircraft accidents in the Royal Air Force of Oman (RAFO). A total of 40 aircraft accidents which occurred in the RAFO between 1980 and 2002 were analysed using HFACS framework. The framework reliability was assessed using Cohen's Kappa and percentage agreement. This study demonstrated that the HFACS system was practicable and reliable for identifying human errors contributing to accidents in the RAFO. The most frequent contributory factors to accidents were at the *unsafe acts of operators* level (46.7%). The lowest frequency was in organisational influences level (9.8%). The system exhibited an acceptable inter- and intra-rater reliability and a high percentage of inter-rater agreement. 14 HFACS categories had a Cohen's Kappa values above 0.40. The percentage agreement figures ranged between 65% and 100%. The HFACS system can be used as a tool in analysing aircraft accidents in the RAFO. The study identified a requirement for crew resource management and aeronautical decision-making programs, professional training for supervisors and efficient resource management for aircraft operations.

IJASM 2013; 57(2): 20-37

Keywords: HFACS, Accidents Analysis, Human Error

### Introduction

It is stated that aviation accidents are the end result of causative factors and hazardous acts of operators. The former contribute to accidents by encouraging human failure. The latter are directly related to the events of accidents. Over the last several decades aircraft have become more reliable, but the inherent reliability of humans has not significantly improved. Human error plays a major role in aviation mishaps. Chappelow stated that "40% of military accidents in the United Kingdom (U.K) are attributed to aircrew human factors, and many of the 17% classified as 'not positively determined' probably have a strong human factors element"(2).

There has been no comprehensive analysis of data on aviation accidents previously in the Royal Air Force of Oman (RAFO). As with all air force

operations, safe operation is of the highest priority and so this study was considered an essential contribution to flight safety. The aim of this research was to discover the most common human factors contributing to RAFO accidents and to look at the applicability and practicability of the use of HFACS in RAFO. In addition, the inter-rater and intra-rater reliability of the 18 individual categories of HFACS framework were examined.

Li and Harris opined that "the role of human errors in aircraft accidents has received a broad scientific discussion in the aviation industry" (9). Aviation accidents do not happen in isolation; they are a result of a sequence of events. The concern

---

Royal Air Force of Oman, P.O. Box: 722  
Postal Code: 111, Muscat, Sultanate of Oman  
Tel: +968 24316535(Y.M. Al-Wardi)  
Fax: +968 24314157  
E-mail: [yalwardi2000@yahoo.com](mailto:yalwardi2000@yahoo.com)

is to understand why a pilot's actions made sense to him prior to the accident.

There are several frameworks and taxonomies that have been developed for describing and analysing human error. Each of these is based on different viewpoints about how human error as a causal factor contributes to the series of events resulting in accidents. Beaubien and Baker reviewed a number of these frameworks and taxonomies(1). The Human Factors Analysis and Classification System (HFACS) adapted from Reason (11) by Wiegmann and Shappell (16) has been found to have some strengths and weaknesses. The core strength point is that the framework is organised in an efficient, hierarchical structure. The HFACS can be adapted to be used in a diversity of industries for safety actions. In addition, it is comparatively comprehensive. Beaubien and Baker discussed the limitations of the HFACS(1). They stated that "the system may be too coarse to identify specific operational problems or to suggest interventions for those problems". The other concern was that "the HFACS taxonomy does not identify the chain of events. As a result, it is difficult to separate causes from effects, even at the most granular level". Nevertheless, the same authors while commenting on taxonomies of human performance and human error such as HFACS have stated that "they are somewhat better suited to identifying the specific problems faced by an individual carrier and for suggesting targeted interventions to those problems".

HFACS has been shown to be useful within the U.S.N as a tool for investigating and analysing the human causes of aircraft accidents (12,13,14). Li and Harris found that HFACS framework can be used to identify the human factors associated with accidents in the Republic of China (R.O.C) Air Force (9). They also observed that the framework is reliable and culture-sensitive. It has

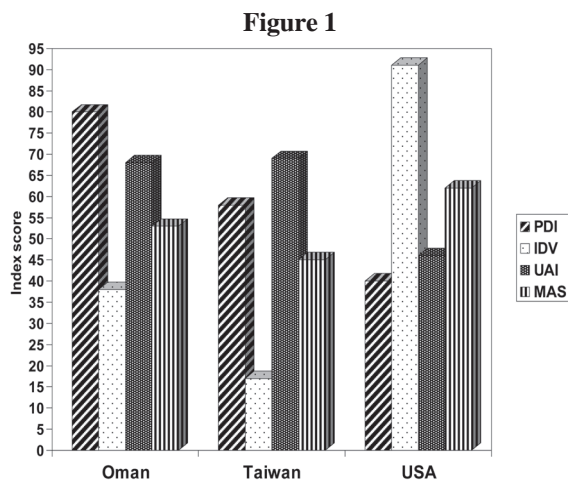
been demonstrated that the HFACS framework is also practicable for use within the civil sector (4, 15).

This study applies the HFACS framework to the existing investigation reports (Board of Inquiry) for the analysis of the human factors causes of aircraft accidents in the RAFO. The Board of Inquiry Team is appointed by the Commander of the Royal Air Force of Oman (CRAFO) and consists of four to six pilots and engineering officers. These officers are experienced in the investigation of aircraft accidents and have the authority to collect all the information related to the accidents and interview all the people concerned.

The HFACS framework was established on Reason's 'Swiss cheese' model (16). The framework relates unsafe acts of the operator with the latent conditions that are present in the system and that combine with environmental factors. HFACS consists of four layers, as represented in Reason's model. However, it provides more detail about human error at each level to be utilised during the investigation or analysis of aircraft accidents. The four levels are subdivided into several categories. Level one (*Unsafe acts of operators*) is where the active failures take place that may lead to accidents. It is classified into four categories: *decision errors, skill-based errors, perceptual errors* and *violation*. Level two (*Preconditions for unsafe acts*) deal with the latent conditions and the more apparent active failures. It also contains categories relating to the operators' condition and the working environment. The categories in this level are: *adverse mental states; adverse physiological states; physical/mental limitation; crew resource management; personal readiness; physical environment; and technological environment*.

The dormant conditions are addressed in levels

three and four. Level three (*Unsafe supervision*) traces the contributory factors for unsafe acts back up the supervisory chain of command. It includes: *inadequate supervision; planned inadequate operations; failed to correct a known problem;* and *supervisory violation*. Level four (*Organisational influences*) identifies the latent conditions that could affect supervisory practices as well as operators. This is categorised into: *resource management, organisational climate* and *organisational process*. (See figure 1).



The reliability of HFACS in this study was assessed by using the Cohen’s Kappa index. It measures the agreement between two raters for qualitative data. It is generally thought to be a robust measure since it takes into account the agreement occurring by chance.

After completing the first coding process by the three independent raters, (an aviation psychologist from U.K; human factors specialist from Taiwan and the author) the inter-rater agreement was assessed. On the completion of second coding of the accidents done only by the author, the intra-rater and inter-rater agreements were measured again. This achieved the aim of examining the intra-rater and inter-rater reliabilities of the HFACS framework both as a whole and as individual categories. The Cohen’s Kappa values

were interpreted using the scale proposed by Landis and Koch (1977) (8). (*The K value range of 0.80-1.00: Almost perfect; 0.60-0.79: Substantial; 0.40-0.59: Moderate; 0.20-0.39: Fair; 0.00-0.19: Slight and <math>d^{\prime}</math>0.00: Poor*) However it has been found that Cohen’s Kappa, in some cases, is unreliable in obtaining reliability between the raters. Gwet has stated that “Kappa does not take in account raters’ sensitivity and specificity” (5). The same author stated that “Kappa becomes unreliable when raters’ agreement is either very small or very high”. Huddleston also, found that Kappa calculation can be misleading when most of the rater’s observations fall in one category (7). To overcome Cohen’s Kappa limitations, the inter- and intra-rater reliabilities were also calculated using simple percentage rate.

The inter-rater reliability of the HFACS framework has been assessed in several studies. Wiegmann et al., in an empirical evaluation demonstrated that the Kappa indices were within ‘Good’ to ‘Excellent’ (17). In analysing U.S civil aviation accidents between 1990 and 1996 the overall Kappa value for the framework was 0.71 which was at the ‘good’ level (15). The HFACS framework had a percentage agreement between raters of between 77% and 83% (16). Li and Harris found that fourteen individual categories had Kappa values of more than 0.60, which signified acceptable agreement. The remaining categories had Kappa values of between 0.40 and 0.59 representing moderate levels of agreement (9). The same authors outlined that there was an acceptable agreement between the raters of between 72.3% and 96.4% when a simple percentage was calculated. Wiegmann and Shappell found that the framework exhibited higher inter-rater reliability in the first and second levels than the third and fourth levels (16), whereas Li and Harris demonstrated that categories at the unsafe

acts of operators' level had the lowest levels of inter-rater reliability (9).

Olsen and Shorrock found that the HFACS adapted by the Australian Defence Force (ADF) was unreliable while using it to assess its suitability as an incidents analysis frameworks at the air traffic control (ATC) section level(10). The study showed that, the result of inter-coder consensus and intra-coder consistency was unacceptable. The percentage agreement level was between 40 and 45% over the study.

## Methods

Data were collected from the reports of the Boards of Inquiry (BOI) of accidents occurring between 1980 and 2002, from the flight safety department in the RAFO. A summary of each report was obtained by the researcher, as the full reports were restricted by confidentiality. The summary was limited to the narrative description of the report and did not include the BOI conclusions. This is to avoid bias during the rating process. The summary was read and 40 were found to have underlying human factors causes. The rest were attributed to unknown causes or mechanical failures. These 40 reports were then studied and analysed in detail. The contributory human factors were coded into the HFACS framework. It is worth mentioning that none of the raters played any role in the investigation of the accidents.

The data analysis was carried out on the 40 accidents identified. The 40 reports were rated by the three raters. The three raters agreed that each of the 18 categories was counted only once per accident to avoid over representation. Also, it was decided that the code was an indicator of the presence of any one of the 18 categories in a given accident report. The rating process took place independently as the aim was to establish the

framework inter-rated reliability. A few weeks after the initial classification of the accidents, the author again coded the reports. The objective was to assess the intra-rater reliability of the HFACS.

An electronic data file was produced using the Statistical Package for the Social Sciences (SPSS) software version 13. For each category in the HFACS framework, a variable was created and assigned with a short label. In order to have an overview of the underlying human factors causes and the distribution of the cause's outcome, the descriptive statistics was a vital stage. This was done by producing a descriptive output of frequencies for all the variables in the form of frequency. This also achieved the aim of establishing the frequency of each level in the HFACS framework.

## Results

A total of 40 RAFO accidents between 1980 and 2002 were analysed by the author twice at a four week interval using the HFACS framework. In these accidents, 122 instances of human error were coded within the HFACS framework during the first analysis and 129 instances the second time. This showed the total human errors identified in the 40 accidents.

The results showed that each level in the HFACS framework exhibited different frequencies of occurrence. Level-1 (*Unsafe acts of operators*) was implicated in the highest number of occurrences on both occasions done by the author. There were 57 (46.7%) occurrences in the first coding and 58 (45%) in the second coding. The second most frequently used was level-2 (*Preconditions for unsafe acts*) with 38 (31.1%) occurrences in the first coding and 41 (31.8%) in the second coding. In the first coding, the lowest frequency was in level-4 (*Organisational influences*) with 12



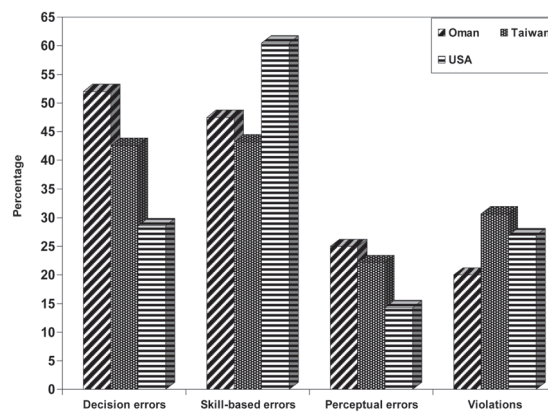
(9.8%) occurrences. Level-3 (*Unsafe supervision*) was involved in 15 (12.3%) occurrences. In the second coding both level-3 (*Unsafe supervision*) and level-4 (*Organisational influences*) showed the same results (11.6%). Because each accident was a consequence of more than just a single factor in the HFACS framework, the percentages quoted add to more than 100% across the whole results. The results were presented as the percentage of accidents associated with at least one casual factor. Also the frequency of occurrence of HFACS categories was obtained for all four levels.

*Unsafe acts of operators* level showed the highest number of occurrences; 57 (46.7%) in the first coding and 58 (45%) in the second coding (both codings done by the author). In this level, the *skill based errors* category demonstrated the highest number of occurrences, 21 instances (52.5%) in the first coding and 18 (47.5%) in the second coding. Incidents in this category included: inappropriate stick and rudder co-ordination, over control of the aircraft, failure to hold an accurate hover, over-reliance on automation, inappropriate recovery from dive, improper calculation of takeoff or landing weight, breakdown in visual scan and failure to operate systems according to procedures. The *decision errors* category presented the highest frequency of occurrence with 21(52.5%) instances in the second analysis and it had the second highest rate in the first analysis of 18(45%) instances. These errors included: the inappropriate decision to fly at low altitude, exceeding the ability of the aircraft, ignoring of warnings and inaccurate or delayed responses to an emergency.

The category of *perceptual errors* exhibited the same frequency of occurrence in both instances of coding the data; that is 10 cases (25%). Factors in this category included: optical illusions, misjudged distance or clearance and disorientation. The

*violations* category showed the lowest frequency of occurrence with eight instances (20%), which was the same in both analyses carried out by the author. These instances included: unauthorised low-altitude flying, inadequate briefing for flight and a failure to comply with orders and instructions. The frequency and percentage of factors implicated in accidents at level-1 (*Unsafe acts of operators*) during both analyses are presented in figure 2.

Figure 2

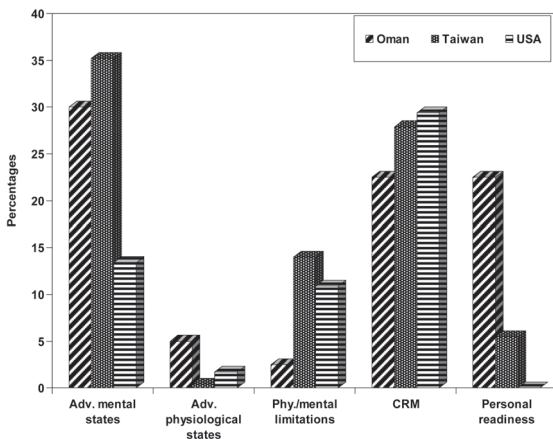


Level-2 of the HFACS framework demonstrated the second highest number of occurrences as contributory factors in accidents. It had 38 (31.1%) in the first analysis and 41 (31.8%) in the second ,where both analyses were done by the author. The *adverse mental states* category showed the highest frequency of causal factors in both cases; (27.5%) and (30%). This included: overconfidence, channelised attention, high workload and distraction. The next most common occurrence was the *personal readiness* category (20% in the first analysis). This was mostly due to inadequate training and poor risk judgment. There were no instances of self-medication, poor diet or overexertion. In the second analysis the frequency of occurrence in the *personal readiness* category was the same as the number of occurrences in the *crew resource management* category (22.5%). The *crew resource management* (CRM) category included issues such as: failure of leadership,

breakdown of communication, insufficient briefing and poor team work. The *physical/mental limitations* category exhibited the same frequency as the CRM category in the first analysis. This included: instances of visual limitations, lack of sensory input and inadequate experience for the complexity of the task. In the second analysis, instances in the *physical/mental limitations* category reduced from six to only one, which could be attributed to the experience gained by the author.

In the *physical environment* category, the causal factors were due to inappropriate reactions to factors in the environment such as poor lighting, terrain and bad weather. The *technological environment* category was involved in only a few accidents. This included equipment failure and head up display (HUD) problems. There were no obvious differences in the frequency of occurrence in the *physical environment* and *technological environment* categories during both analyses. The lowest number of occurrences was exhibited by the *adverse physiological states* category which included physical fatigue. (See figure 3).

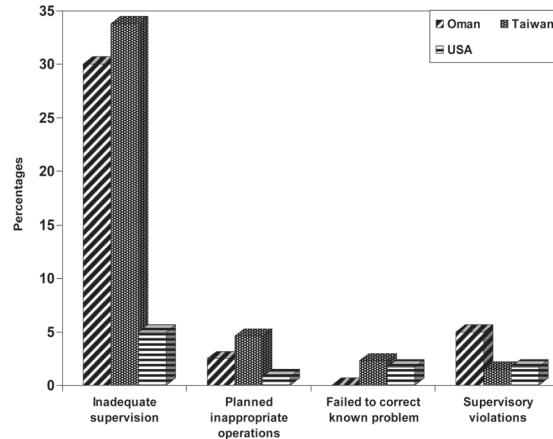
Figure 3



The *unsafe supervision* level of the HFACS framework contained 15 instances in both analyses carried out by the author. There were 122 instances of human error coded during the first analysis and 129 instances in the second. So, *unsafe supervision*

level accounted for (12.3%) and (11.6%) respectively as casual factors implicated in aircraft accidents. The *inadequate supervision* category had the highest frequency of occurrence in this level as a contributory factor in aircraft accidents with 12 (30%) occurrences. This encompassed: a failure to check qualifications and track performance, not providing sufficient information and a failure to provide professional guidance. In addition, there was a failure to provide adequate rest periods and a lack of authority and lack of responsibility. *Planned inadequate operations, a failure to correct a known problem* and *supervisory violations* categories were involved in very few accidents. Each one was implicated in only one instance in the first analysis. This included: failure to provide adequate supervision, a failure to identify risky behaviour and authorised unqualified crew for flight. There were no fundamental differences in the second analysis. (See Figure 4).

Figure 4



The lowest frequency of occurrence in the first analysis done by the author was at the *organisational influences* level. There were 12 (9.8%) instances contributing to accidents. In the second analysis done by the author the *organisational influences* level exhibited the same number of occurrence to that exhibited at the *unsafe supervision* level. The *organisational process* category, which included: poor risk management

programs, failures in the checking process, procedure problems and improper mission scheduling; was the most common occurrence in accidents in the first analysis. The next most common occurrence was in the *resource management* category. Training of human resources was the main topic in this category. Providing unsuitable equipment and lack of funding were less frequent causes. In the second analysis both the *resource management* and *organisational process* categories showed identical results. The *organisational climate* category, in both analyses, was only involved once as a contributory factor in accidents. This was related to poor communication. (See Figure 5).

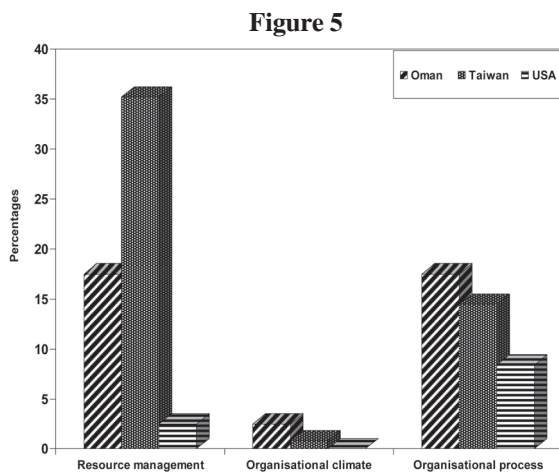


Table 1 shows the inter-rater reliabilities (Cohen’s Kappa and percentage agreement) between the first rater (the author) and the second rater (human factors specialist) of the HFACS framework categories after both first and second coding done by the author.

In table 1, four HFACS categories had a Cohen’s Kappa of below 0.40 and four categories were constant. The rest had Cohen’s Kappa values above 0.40 and one HFACS category had perfect agreement. After the second coding done only by the author, the number of Cohen’s Kappa values below 0.40 reduced to three. Also, there was a

considerable improvement in the Cohen’s Kappa values across the HFACS categories. This could explain the issue about the HFACS practice and learning.

Table 2 shows the results between the first rater (the author) and the third rater (aviation psychologist) of the HFACS framework categories after both first and second coding done by the author.

It can be seen from table 2 that seven HFACS categories had Cohen’s Kappa values below 0.40 and only one category was constant. The other categories had Cohen’s Kappa values above 0.40 and one category had a Kappa value of one. The Cohen’s Kappa less than 0.40 reduced to two after the second coding carried out by the author. Also the Cohen’s Kappa values improved significantly for the entire HFACS framework. This result can be attributed to the learning gained by the author in the process of two codings.

Table 3 shows the results between the second rater (human factors specialist) and the third rater (aviation psychologist) of the HFACS framework categories.

In table 3, three HFACS categories had a Cohen’s Kappa of below 0.40 and one category was constant. The rest had Cohen’s Kappa values above 0.40 and five HFACS categories had perfect agreement. The result might be influenced by the experience and knowledge of HFACS that the two raters have (human factors specialist and aviation psychologist).

The inter-rater reliabilities of the HFACS framework categories were ranked in terms of the increasing rate of percentage agreement. From table 1 and 2, the percentage agreement figures ranged between 65% and 100% for the first coding and between almost 75% and 100% for the second

**Table I the Inter-rater reliability (Cohen’s kappa and percentage agreement), for the 1<sup>st</sup> (The author) and the 2<sup>nd</sup> rater (Human Factors Specialist) (both codings done by the author) of HFACS Categories (Ranked in order of increasing agreement).**

HFACS Categories 1 <sup>st</sup> Coding	HFA CS level	Inter-rater Reliability 1 <sup>st</sup> rater & 2 <sup>nd</sup> Rater		HFACS Categories 2 <sup>nd</sup> Coding	HFA CS level	Inter-rater Reliability 1 <sup>st</sup> rater & 2 <sup>nd</sup> Rater	
		Cohen’s Kappa (significance)	Percentage agreement			Cohen’s Kappa (significance)	Percentage agreement
Inadequate supervision	3	0.239 (0.121)	65.0%	Inadequate supervision	3	0.348 (0.024)	70.0%
Decision errors	1	0.442 (0.005)	72.5%	Skill-based errors	1	0.501 (0.001)	75.0%
Skill-based errors	1	0.499 (0.002)	75.0%	Perceptual errors	1	0.429 (0.006)	80.0%
Crew resource management	2	0.174 (0.268)	77.5%	Crew resource management	2	0.533 (0.001)	85.0%
Perceptual errors	1	0.429 (0.006)	80.0%	Adverse mental states	2	0.643 (0.000)	85.0%
Adverse mental states	2	0.573 (0.000)	82.5%	Decision errors	1	0.702 (0.000)	85.0%
Physical/mental limitations	2	Constant	85.0%	Technological environment	2	0.286 (0.053)	90.0%
Organisational process	4	0.474 (0.003)	87.5%	Violations	1	0.737 (0.000)	90.0%
Technological environment	2	-0.053 (0.739)	90.0%	Adverse physiological states	2	-0.034 (0.816)	92.5%
Resource management	4	0.660 (0.000)	90.0%	Supervisory violations	3	Constant	95.0%
Violations	1	0.737 (0.000)	90.0%	Physical environment	2	0.643 (0.000)	95.0%
Adverse physiological states	2	-0.026 (0.871)	95.0%	Organisational process	4	0.805 (0.000)	95.0%
Physical environment	2	0.643 (0.000)	95.0%	Resource management	4	0.844 (0.000)	95.0%
Personal readiness	2	0.857 (0.000)	95.0%	Physical/mental limitations	2	Constant	97.5%
Planned inadequate operations	3	Constant	97.5%	Planned inadequate operations	3	Constant	97.5%



HFACS Categories 1 <sup>st</sup> Coding	HFA CS level	Inter-rater Reliability 1 <sup>st</sup> rater & 2 <sup>nd</sup> Rater		HFACS Categories 2 <sup>nd</sup> Coding	HFA CS level	Inter-rater Reliability 1 <sup>st</sup> rater & 2 <sup>nd</sup> Rater	
		Cohen's Kappa (significance)	Percentage agreement			Cohen's Kappa (significance)	Percentage agreement
Failed to correct a known problem	3	Constant	97.5%	Personal readiness	2	0.931 (0.000)	97.5%
Supervisory violations	3	Constant	97.5%	Failed to correct a known problem	3	Constant	100%
Organisational climate	4	1.000 (0.000)	100%	Organisational climate	4	1.000 (0.000)	100%

**Table 2 the inter-rater reliability (Cohen's Kappa and percentage agreement), for the 1<sup>st</sup> (The author) and the 3<sup>rd</sup> (Aviation Psychologist) rater (Both codings done by the author) of HFACS categories (Ranked in order of increasing agreement)..**

HFACS Categories 1 <sup>st</sup> Coding	HFA CS level	Inter-rater Reliability 1 <sup>st</sup> rater & 3 <sup>rd</sup> Rater		HFACS Categories 2 <sup>nd</sup> Coding	HFA CS level	Inter-rater Reliability 1 <sup>st</sup> rater & 3 <sup>rd</sup> Rater	
		Cohen's Kappa (significance)	Percentage agreement			Cohen's Kappa (significance)	Percentage agreement
Decision errors	1	0.293 (0.064)	65.0%	Skill-based errors	1	0.501 (0.001)	75.0%
Inadequate supervision	3	0.425 (0.006)	72.5%	Inadequate supervision	3	0.521 (0.001)	77.5%
Skill-based errors	1	0.499 (0.002)	75.0%	Organisational process	4	0.654 (0.000)	80.0%
Physical/mental limitations	2	-0.081 (0.542)	80.0%	Perceptual errors	1	0.600 (0.000)	85.0%
Crew resource management	2	0.216 (0.173)	80.0%	Crew resource management	2	0.593 (0.000)	87.5%
Organisational process	4	0.358 (0.023)	82.5%	Decision errors	1 (0.000)	0.751	87.5%
Adverse mental states	2	0.573 (0.000)	82.5%	Violations	1	0.714 (0.000)	90.0%
Perceptual errors	1	0.600 (0.000)	85.0%	Personal readiness	2	0.734 (0.000)	90.0%
Personal readiness	2	0.658 (0.000)	87.5%	Adversement states	2	0.762 (0.000)	90.0%
Technological environment	2	0.398 (0.002)	87.5%	Adverse physiological states	2	-0.034 (0.816)	92.5%

HFACS Categories 1 <sup>st</sup> Coding	HFA CS level	Inter-rater Reliability 1 <sup>st</sup> rater & 3 <sup>rd</sup> Rater		HFACS Categories 2 <sup>nd</sup> Coding	HFA CS level	Inter-rater Reliability 1 <sup>st</sup> rater & 3 <sup>rd</sup> Rater	
		Cohen's Kappa (significance)	Percentage agreement			Cohen's Kappa (significance)	Percentage agreement
Planned inadequate operations	3	-0.039 (0.773)	90.0%	Physical/mental limitations	2	-0.034 (0.816)	92.5%
Resource management	4	0.600 (0.000)	90.0%	Technological environment	2	0.688 (0.000)	92.5%
Violations	1	0.714 (0.000)	90.0%	Supervisory violations	3	0.474 (0.003)	95.0%
Supervisory violations	3	-0.034 (0.816)	92.5%	Planned inadequate operations	3	0.481 (0.000)	95.5%
Adverse physiological states	2	-0.026 (0.871)	95.0%	Physical environment	2	0.643 (0.000)	95.5%
Physical environment	2 (0.000)	0.643	95.0% management	Resource	4 (0.000)	0.844	95.5%
Failed to correct a known problem	3	Constant	97.5%	Failed to correct a known problem	3	Constant	100%
Organization climate	4	1.000 (0.000)	100%	Organisational climate	4	1.000 (0.000)	100%

coding. It is obvious that, after the second analysis done by the author, the simple percentage agreement figures were improved significantly across the whole HFACS categories. From table III, the percentage agreement figures ranged between 87.5% and 100%. Overall, the inadequate supervision category showed the lowest inter-rater reliability.

In general, level-1 (*Unsafe acts of operators*) in the HFACS framework exhibited low inter-rater reliabilities whereas, level-4 (*Organisational influences*) had substantial inter-rater reliabilities. Looking closely at three tables, level-2 (*Preconditions for unsafe acts*) demonstrated moderate agreement, which was

lower than the level of agreement shown by level-3 (*Unsafe supervision*). This continued to be the same even after the second coding done by the author, but there was an improvement in the inter-rater reliabilities across the HFACS framework. From table I, the average percentage agreement between the author and the human factors specialist was 87.4% and 90.3% respectively. On the other hand, it was 86% and 90% between the author and the aviation psychologist. The improvement in reliabilities after the second coding done by the author could be attributed to the learning process. From table III, the average percentage agreement between the human factors specialist and the aviation psychologist was 93.8%

**Table 3** the inter-rate reliability (Cohen’s kappa and percentage agreement), for the human factors specialist (2<sup>nd</sup> rater) and aviation psychologist (3<sup>rd</sup> rater) of HFACS categories (ranked in order of increasing agreement).

HFACS categories	HFACS level	Inter-rater reliability 2 <sup>nd</sup> Rater & 3 <sup>rd</sup> Rater	
		Cohen’s Kappa (Significance)	Percentage Agreement
Technological environment	2	0.398 (0.002)	87.5%
Inadequate supervision	3	0.742 (0.000)	87.5%
Decision errors	1	0.848 (0.000)	92.5%
Planned inadequate operations	3	0.000	92.5%
Personal readiness	2	0.000	95.0%
Physical/Mental limitations	2	0.000	95.0%
Supervisory violations	3	0.000	95.0%
Adverse mental states	2	0.881 (0.000)	95.0%
Perceptual errors	1	0.857 (0.000)	95.0%
Violations	1	0.857 (0.000)	95.0%
Crew resource management	2	0.908 (0.000)	97.5%
Organisational process	4	0.805 (0.000)	100%
Resource management	4	1.000 (0.000)	100%
Skill-based errors	1	1.000 (0.000)	100%
Failed to correct a known problem	3	Constant	100%

The intra-rater reliability was assessed using Cohen’s Kappa and simple percentage rate (See table 3). There were no major differences in frequency of occurrence of HFACS categories implicated in accidents between the first and second coding by the author except for *physical/mental limitations* category. The Cohen’s Kappa coefficient ranged between zero and one. There

were three HFACS categories that had Kappa values of less than 0.40 and one category was a constant. The *decision errors* category showed moderate Kappa coefficient (0.552). Three HFACS categories exhibited Cohen’s Kappa values of between 0.61 and 0.80 which was regarded as substantial agreement. Ten HFACS categories exceeded a Cohen’s Kappa value of 0.80 which is

**Table 4** the frequency of occurrence (both codings done by the author) and intra-rater reliability (Cohen's kappa and percentage agreement), of HFACS Categories (Ranked in order of increasing agreement).

HFACS categories	HFACS Level	Frequency of occurrence 1 <sup>st</sup> coding	Frequency of occurrence 2 <sup>nd</sup> coding	Intra-rater Reliability	
				Cohen's kappa (Significance)	Percentage agreement
Decision errors	1	18	21	0.552 (0.000)	77.5%
Physical/Mental limitations	2	6	1	0.254 (0.016)	87.5%
Supervisory violations	3	1	2	-0.034 (0.816)	92.5%
Organisational process	4	6	7	0.725 (0.000)	92.5%
Crew resource management	2	6	9	0.756 (0.000)	92.5%
Adverse mental states	2	11	12	0.817 (0.000)	92.5%
Planned inadequate operations'	3	1	1	-0.026 (0.871)	95.0%
Technological environment	2	2	4	0.643 (0.000)	95.0%
Resource management	4	5	7	0.805 (0.000)	95.0%
Inadequate supervision	3	12	12	0.881 (0.000)	95.0%
Skill-based errors	1	21	19	0.900 (0.000)	95.0%
Failed to correct a known problem	3	1	0	Constant	97.5%
Adverse physiological states	2	1	2	0.655 (0.000)	97.5%
Personal readiness	2	8	9	0.925	97.5%
Perceptual errors	1	10	10	1.000 (0.000)	100%
Violations	1	8	8	1.000 (0.000)	100%
Physical environment	2	4	4	1.000 (0.000)	100%
Organisational climate	4	1	1	1.000 (0.000)	100%

considered as almost perfect agreement. The intra-rater reliabilities of the HFACS framework categories were ranked in terms of the increasing rate of percentage agreement. The *decision errors* category had the lowest percentage rate of agreement. The percentage rate of agreement figures ranged between 77.5% and 100% where four categories had 100% agreement. The average percentage agreement was 94.6%. This indicates extraordinarily high rate of intra-rater reliability of the HFACS framework. Although one category at level one showed the lowest percentage agreement, two categories demonstrated very high agreement. This result might be influenced by the relatively short time between the two codings and the author's familiarity with the relatively low number of cases. The other HFACS framework levels did not show any specific pattern for the intra-rater reliability.

## Discussion

This study supported the results of previous studies found in the literature. Studies have shown that the *unsafe acts of operators* level exhibited the highest rate as a causal factor contributing to the aircraft accidents. The *skill-based errors* category was the most common error at this level (3, 9). The *perceptual errors* and *violations* categories were associated with a low rate of occurrence as contributory factors implicated in the accidents (4, 15). In this study the *unsafe acts of operators* level was associated with the highest frequency of contributory factors in aircraft accidents. In the first analysis it accounted for 57 (46.7%) instances and 58 (45%) in the second analysis. *Skill-based errors* had the highest frequency the first time (52.5%) and it was the second most common in the second analysis (47.5%). *Decision errors* showed the highest frequency (52.5%) in the second analysis and it had the second highest frequency of occurrence in

the first analysis (45%). *Skill-based errors* and *decision errors* exhibited an almost identical frequency in both analyses.

The category of *perceptual errors* exhibited the same frequency of occurrence in both analyses done by the author, which was 10 cases (25%). *Violations* showed the lowest frequency of occurrence with eight instances (20%) which was the same in both analysis. The low representation of the last two categories may be attributed to the sample size analysed. Another explanation may be that these issues are not being looked for during the aircraft accident investigations carried out in the RAFO. It is essential to identify practices which are essentially violations, but over time have become common practices. *Violations* can be reduced further by enforcing good rules and having clear procedures.

This research demonstrated that *preconditions for unsafe acts* showed the second highest frequency as contributory factors in accidents. This is no different from previous reports in the literature which explains the importance of operators' conditions, operators' practices and the environment. Research has shown that the *preconditions for unsafe acts* level exhibited the second highest rate of occurrence (4, 9, 15). Li and Harris found that the *adverse mental states* category had the highest rate of occurrence(9), whereas Wiegmann and Shappell stated that Crew Resource Management (CRM) breakdown was involved in the highest rate of accidents(15). The former study was carried out in military aviation in contrast to the latter which was in civil aviation, involving more instances of multiple crews. Another explanation was related to the different cultures in which each study was conducted. In this study *preconditions for unsafe acts* had a 31.1% occurrence in the first analysis and 31.8% in the



second; (both analyses done by the author). The *adverse mental states* category showed the highest frequency of causal factors in both cases; 27.5% and 30%. This high rate of accidents can be reduced by tackling them at various levels. Flying an aircraft is a challenging task, so pilot selection, training and professional development are critical issues.

Communication is necessary in military aviation between the pilots themselves, the leader of the squadron and the air traffic controller. It is essential to exchange information and to co-ordinate with the mission leader in order to achieve safe operations. The *crew resource management* (CRM) category was the second highest category. It had a similar rate to the *personal readiness* category (22.5%). This makes the requirement for a structured CRM training program in the aviation industry paramount. This study illustrated the need for a CRM training program to be implemented and the effectiveness of this program should be assessed by follow up audits. Resources might be invested in both aeronautical decision-making and CRM programs which should have practical training. The other categories in level-2 of the HFACS framework: (*adverse physiological states, physical/mental limitation, physical environment and technological environment*) demonstrated low frequencies of causal factors in aircraft accidents. This could be a result of being overlooked during the accident investigation or they were simply not considered.

The *unsafe supervision* is the third level in the HFACS framework. In analysing accidents in U.S civil aviation between 1990 and 1996, Wiegmann and Shappell found that causal factors at this level were involved in 9.2% of accidents(15). *Unsafe supervision* contributed to 12.5% of accidents in R.O.C Air Force (9). *Inadequate supervision* was the most common category,

accounting for 33.8% of instances. This was also seen in civil aircraft accidents that took place in India where *inadequate supervision* had the highest amongst the categories at the *unsafe supervision* level (4).

Analysing the data for this study, the *unsafe supervision* level accounted for 12.3% and 11.6% of occurrences as a casual factor implicated in aircraft accidents in each analysis done by the author which are similar to that found in R.O.C Air Force (9). The supervisors' roles and responsibilities in military aviation are, usually, clearly defined. The study found that *inadequate supervision* category had the highest frequency (30%) of occurrence. The supervisors' behaviour, directly or indirectly, affects the flight crew actions. Therefore, it is essential that supervisors carry out their tasks efficiently and professionally. The supervisors have to practice their authority and responsibility, so safe operations can be performed. The supervisors in the aviation industry need to get adequate and structured training in order to carry out their tasks effectively and efficiently. *Planned inadequate operations, a failure to correct a known problem* and the *supervisory violations* categories were involved in very few accidents. However, this could be attributed to the way the accident investigations were conducted as there is no clear guidance for these issues.

This study has demonstrated that *organisational influences* showed the lowest frequency of occurrence in the first analysis. In the second analysis, the frequency of occurrence at the organisational influences level was the same as that for the unsafe supervision level(both analys done by the author). In reviewing the literature, the *organisational influences* level represented the lowest rate as a contributory factor in aircraft accidents except in the study reported by Gaur (4).

He found that *organisational influences* contributed to 52.1% of the accidents. Within that level, *organisational process* represented the highest rate (41.7%). Accidents occurring in the R.O.C. Air Force between 1978 and 2002 showed that 15% of the instances were related to the *organisational influences*, where *resource management* category was the most frequent (9). In U.S civil aviation between 1990 and 1996, *organisational influences* were a causal factor in 10.9% of the accidents. *Organisational process* was the most commonly cited category at this level. (15). It is obvious that in civil aviation the *organisational process* category was the most common contributory factor whereas the *resource management* category was the most frequent category in military aviation. This might be as a result of the different types of operation carried out by each sector.

This research confirmed that *organisational process* category was the most common occurrence in accidents in the first analysis (15%). In the second analysis both the *resource management* and *organisational process* categories showed an identical rate of occurrence (17.5%). A lack of training in human resources was the main topic in the *resource management* category. Providing unsuitable equipment and lack of funding were less frequent causes. The *organisational climate* category, in both analyses, was only involved once as a contributory factor in accidents. The low frequency in both the first and the second analysis carried out by the author may be due to the low number of accident reports analysed in this study. The other explanation is the possibility that it was not assessed during the accident investigations. Nevertheless, the results demonstrated that *organisational process* and *resource management* have a similar frequency of occurrence as contributory factors in aircraft

accidents. It is not easy to trace the latent conditions at the organisation level. The factors such as: poor risk management programs, failures in the checking process, procedure problems and improper mission scheduling played a major role in accidents. It is crucial to explore the weaknesses at the organisational level when implementing strategic remedies.

In this study, the HFACS framework exhibited an acceptable level of agreement among the raters coding the data, even though these raters were from different cultures (the author from Oman, human factors specialist from Taiwan and aviation psychologist from U.K) and had different experience in using the system. This indicates the applicability and practicability of the HFACS framework. There were 11 individual categories between the author and the human factors specialist which had Kappa values of between 0.40 and 1.00. This indicates moderate to substantial agreement between the raters. Fifteen individual categories had Kappa values of between 0.40 and 1.00 between the author and the aviation psychologist which again signifies moderate to substantial agreement. The human factors specialist and aviation psychologist result showed 13 individual categories which had moderate to substantial agreement; Kappa values of above 0.40. The overall percentage agreement of the HFACS system was found to be between 86.7% and 93.8%. This shows high agreement between the raters.

The reliability was assessed by using the Cohen's Kappa index and also calculating the percentage rate of agreement. This gave some differences between Kappa values and percentage of agreement in some of the HFACS categories. For example, the *adverse physiological states* category had a Kappa coefficient of 0.000, but had percentage agreement of 95%. *Failed to correct*

a known problem category had a Kappa value returning a constant but it had 97.5% percentage agreement in the first analysis and 100% in the second analysis where both analysis carried out by the author. On the other hand, the Kappa index and percentage of agreement were in strong agreement in some cases. The *organisational climate* category had a Kappa coefficient of 1.000 and a percentage agreement of 100% between all raters. In general there was some discrepancy between the Kappa index and percentage agreement, especially in those categories with low frequency of occurrences. The Cohen's Kappa index can be unreliable when most of the instances fall in one category. Also, Kappa values can be affected when raters' agreement is either very small or very high. The percentage agreement figure was calculated to overcome Cohen's Kappa limitations.

Overall, the *inadequate supervision* category showed the lowest inter-rater reliability. Also, the *skill-based errors* category, *decision errors* category and CRM category showed low inter-rater reliabilities. Harris stated that "categories that require a high degree of high presumption when coding the data will show lower levels of agreement"(6). In general, level-1 (*Unsafe acts of operators*) in the HFACS framework exhibited lower inter-rater reliabilities whereas, level-4 (*Organisational influences*) had substantial inter-rater reliability. On the other hand, level-2 (*Preconditions for unsafe acts*) demonstrated moderate agreement, which was lower than the level of agreement shown by level-3 (*Unsafe supervision*). This supports the results of Li and Harris (9) but is contrary to the results of Wiegmann and Shappell (16).

Also, the intra-rater reliability of the HFACS framework was assessed using Cohen's Kappa and simple percentage rate of agreement. Overall, it demonstrated acceptable intra-rater reliability. There

were three HFACS categories that had Kappa values of less than 0.40 and one category that was a constant. The *decision errors* category showed moderate Kappa coefficient (0.552). Three HFACS categories exhibited Cohen's Kappa values of between 0.61 and 0.80, which was regarded as substantial agreement. Ten HFACS categories exceeded a Cohen's Kappa value of 0.80, which can be considered as almost perfect agreement. The HFACS framework had a high intra-rater percentage agreement (94.6%). The *decision errors* category had the lowest percentage rate of agreement. Although one category (*decision errors*) at level one showed the lowest percentage agreement, the rest (*skill-based errors*, *perceptual errors* and *violations*) demonstrated very high agreement. The other HFACS framework levels did not show any specific pattern for their intra-rater reliability. It is difficult to comment on these results, but the high intra-rater reliability of the HFACS framework is evident and can not be dismissed. Olsen and Shorrock (10) found that intra-rater percentage agreement was higher than the inter-rater one.

This study encountered various limitations. The sample size was relatively small which could affect the final result of the study. The rating of the accidents was based on the report summary which was created by the author. This might lead to increase in the inter-rater scores in spite of the ratings being performed independently. Because of the time concern; A few weeks after the initial classification of the accidents, the author again coded the reports. The process influenced the achievement of high level of intra-rater reliability. This was reinforced by the use of the summaries of the BOI reports.

## Conclusions

This study established that the HFACS system

is practicable and applicable for identifying the underlying human factors implicated in aircraft accidents in the RAFO. This conclusion should be treated with caution due to the low number of cases reviewed and the use of summary information. The study showed that using HFACS to analyse accidents, *unsafe acts of operators* were identified as the most frequent causal factor contributing to accidents; however the analysis did not permit this factor to be determined as the root cause. This was followed by the *preconditions for unsafe acts*. This study has demonstrated that *organisational influences* and *unsafe supervision* level accounted for the lowest rate implicated in aircraft accidents. The study pointed to the crucial areas of human factors and errors that required further exploration to support flight safety in the aviation industry. The identification of human factors problems in accidents will help in developing effective prevention strategies in order to promote safety in aviation industry. The HFACS showed an accepted level of inter-rater reliability at individual categories level. In addition the intra-rater reliability of the framework was evident. This suggested that the HFACS framework is a practical tool in identifying human error.

## References

1. Beaubien JM, Baker DP. A review of selected aviation human factors taxonomies, accident/incident reporting system and data collection tools. *International Journal of Applied Aviation Studies* 2002; 2(2):11-36.
2. Chappelow JW. Error and accidents. In: Rainford DJ, Gradwell DP, eds. *Ernsting's Aviation Medicine*. 4<sup>th</sup> ed. London: Hodder Arnold; 2006:349-357.
3. Dambier M, Hinkelbein J. Analysis of 2004 German General Aviation Aircraft Accidents According to the HFACS Model. *Air Medical Journal* 2006; Nov-Dec;25(6): 265-269.
4. Gaur D. *Human Factors Analysis and Classification System* applied to civil aircraft accidents in India. *Aviation, Space, and Environmental Medicine* 2005; 76(5):501-5.
5. Gwet K. Inter rater reliability: Dependency on trait prevalence and marginal homogeneity. *Statistical Methods for Inter-rater Reliability Assessment* 2002b; 1:1-9. Retrieved 01 July 2006 from [http://www.stataxis.com/files/articles/inter\\_rater\\_reliability\\_dependency.pdf](http://www.stataxis.com/files/articles/inter_rater_reliability_dependency.pdf).
6. Harris D. The quantitative analysis of qualitative data: Methodological issues in the derivation of trends from incident reports. In: McDonald N, Johnston R, Fuller R, eds. *Applications of Psychology to the Aviation System*. Vol.1. England: Ashgate publishing, Aldershot; 1995:117-122.
7. Huddleston JA. *An Evaluation of the Training Effectiveness of a Low-fidelity, Multi-player Simulator for Air Combat Training* (Dissertation). England: Cranfield University; (2003).
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-174.
9. Li WC, Harris D. HFACS analysis of ROC Air Force aviation accidents: reliability analysis and cross-cultural comparison. *International Journal of Applied Aviation Studies* 2005; 5(1):65-81.
10. Olsen NS, Shorrock ST. Evaluation of the HFACS-ADF safety classification system: Inter-coder consensus and intra-coder consistency. *Accident Analysis and Prevention Journal* 2010; 42(2): 437-444.
11. Reason J. *Human Error*. New York: Cambridge University, 1990.
12. Shappell SA, Wiegmann DA. Is proficiency eroding among U.S. naval aircrews? A quantitative analysis using the human factors analysis and classification system. In: *Proceedings of the 44<sup>th</sup> Annual Meeting of the Human Factors and Ergonomics Society*. Vol. 4. San Diego, California; 2000: 345-348.
13. Shappell SA, Wiegmann DA, Fraser JR, Gregory G, Kinsey P, Squier H. Beyond mishap rates: A human factors analysis of U.S Navy/Marine Corps TACAIR and rotary wing mishaps using HFACS. *Aviation, Space and Environmental Medicine* 1999; 70(4): 416-417.

14. United States Naval Safety Centre. Human factors analysis and classification system (HFACS): A human error approach to accident investigation, OPNAV 3750.6R November 2001. Appendix O. Retrieved 15 June 2006 from <http://www.safetycenter.navy.mil/aviation/aeromedical/downloads/HFACSPrimer.doc>.
15. Wiegmann DA, Shappell SA. Human error analysis of commercial aviation accidents: application of the Human Factors Analysis and Classification System. *Aviation, Space, and Environmental Medicine* 2001; 72(11):1006-16.
16. Wiegmann DA, Shappell SA. A Human error approach to aviation accident analysis: The Human Factors Analysis and Classification System. Aldershot, England: Ashgate, 2003.
17. Wiegmann DA, Shappell SA, Pape A. A human factors analysis of aviation accident data- An empirical evaluation of the HFACS framework. *Aviation, Space and Environmental Medicine* 2000; 71(3):328.