Using Existing HEI Techniques to Predict Pilot Error: A Comparison of SHERPA, HAZOP and HEIST

Paul Salmon[1], Neville A.Stanton[1], Mark S. Young[1],Don Harris[2], Jason Demagalski[2], Andrew Marshall[3], Thomas Waldman[4], and Sidney Dekker[5]

[1]Brunel University, Department of Design, Egham, Surrey, TW20 OJZ, UK, Paul.salmon@brunel.ac.uk
[2]Cranfield University, Human Factors Group, School of Aviation, Cranfield, Bedford, MK43 OAL, UK
[3]Marshall Associates, London, UK,
[4]University of Limerick, Ireland,
[5]Linkoping University, Sweden

Abstract

At the moment, there appears to be no human error identification (HEI) techniques developed specifically for use in aviation. Similarly, there appears to be very little research concerning the prediction of pilot error in the cockpit. This paper investigates the potential use of existing HEI methods for predicting pilot error and describes a comparative study of three existing HEI techniques, SHERPA, HAZOP and HEIST when used to predict potential pilot error on an aviation landing task using the 'autoland' system. The study aims to demonstrate that existing HEI methods developed for use in highly complex systems, such as nuclear power plants and chemical processing plants, can be used effectively in an aviation context.

Introduction to HEI

Human error in high risk, complex systems, is a problem of great concern to human factor's professionals. When committed by commercial airplane pilots or control room operators in a nuclear power plant, hundreds of lives can potentially be put at great risk. The prediction of human error in these complex systems has therefore been investigated extensively over the past three decades, mainly in response to the increasing number of fatal accidents in highly complex systems attributed to human error, such as Three Mile Island, Bhopal and Chernobyl. As a result of this research, an abundance of human error identification (HEI) techniques were developed throughout the 1980's. Techniques such as THERP, Human Error HAZOP, SHERPA, PHECA and CADA were developed specifically to identify potential human error in high-risk complex systems such as nuclear power plants and chemical processing plants.

HEI in Aviation

It is apparent that the major cause of all aviation accidents is pilot or human error (McFadden and Towell 1999). Recent studies suggest that human error has been identified as the source of at least 60% of the incidents that occur in commercial aviation (McFadden and Towell 1999). Furthermore, studies also suggest that 70% of all aviation accidents are classified as pilot error (BASE 1997, McFadden 1993). When this is coupled with the growing number of high profile aviation disasters involving a large loss of life that are attributed to pilot error (e.g. Tenerife, Mont Saint Odile), it is surprising that there appears to be very little in the way of published research concerning human error identification in aviation. Even more surprising, perhaps, is the apparent lack of proven HEI methods developed specifically for aviation. Indeed, it appears that at the moment, valid and reliable methods for predicting errors on modern day flight decks simply do not exist.

The purpose of this study is to investigate the use of three existing HEI methods, SHERPA, HAZOP and HEIST for their use in predicting potential errors that may occur in a given flight task. The study also aims to compare the three methods in terms of efficiency in their ability to predict the correct errors. Each of the methods was applied to a Hierarchical task analysis (Annett, Duncan, and Stammers 1971) of the landing task, 'Landing at New Orleans using the autoland system'. The errors predicted by the methods were then compared to actual error data, supplied by Cranfield University. A brief description of each of the methods is given below:

Systematic Human Error Reduction and Prediction Approach (SHERPA)

SHERPA (Embrey 1986) uses hierarchical Task Analysis (HTA) (Annett, Duncan, and Stammers 1971) together with an error taxonomy to identify credible errors associated with a sequence of human activity. The SHERPA technique works by indicating which error modes are credible for each task step in turn, based upon an analysis of work activity. This indication is based upon the judgement of the analyst. SHERPA is

Task Step	Error mode	Description	Consequence	Recovery	P	C	Remedial measures
3.2.2	A3	Pilot turns the Speed/MACH selector knob the wrong way	The wrong airspeed is entered and the plane speeds up instead of slowing down	3.2.1	M	M	- Clearer control labelling - Auditory signal informing increase/decrease
3.2.2	A6	The pilot dials in the desired airspeed using the wrong control knob i.e. the heading knob	The auto-pilot will attempt to switch course to the speed value entered causing the plane to leave the glideslope	Immediate	M	H	-Improved control labelling -improved separation of controls

Table 1 – Extract of SHERPA analysis

conducted on each bottom level task step taken from the HTA. Using subjective judgement, the analyst uses the SHERPA human error taxonomy to classify each task step into one of the five following behaviour types:

- **Action** – e.g. dialling in airspeed, moving the flap lever
- **Retrieval** – e.g. retrieving altitude information from the Primary Flight Display
- **Check** – e.g. making a procedural check
- **Selection** – e.g. choosing to set flaps to level 2 instead of level 1
- **Information communication** – e.g. talking to air traffic control or co-pilot

The analyst then uses the taxonomy and domain expertise to determine any credible error modes for the task in question. For each credible error (i.e. those judged by the analyst to be possible) the analyst should give a description of the form that the error would take, such as, 'pilot dials in wrong airspeed'. Next, the analyst has to determine any consequences associated with the error and any error recovery steps that would need to be taken in event of the error. Finally, ordinal probability (Low, medium or high), criticality (Low, medium or high) and any potential design remedies (i.e. how the interface design could be modified to eradicate the error) are recorded. The main strengths of the SHERPA method are that it provides a structured and comprehensive approach to error prediction, gives an exhaustive and detailed analysis of potential errors and also the SHERPA error taxonomy prompts the analyst for any potential errors. Furthermore, a number of studies have shown encouraging validity and reliability data for the SHERPA technique. SHERPA's weaknesses include being both tedious and time consuming to perform and also the fact that SHERPA does not consider the cognitive components of the error mechanisms. The methods consistency when used by different analysts can also be questioned. Table 1 shows an extract of the SHERPA analysis performed in this study.

Human Error Hazard and Operability Study (HAZOP)

HAZOP (Kletz 1974) is a well-established engineering approach that was developed in the late 1960's by ICI (Swann and Preston 1995) for use in process design audit and engineering risk assessment (Kirwan 1992a). Originally applied to engineering diagrams (Kirwan and Ainsworth 1992) the HAZOP technique involves the analyst applying guidewords, such as Not done, More than or Later than, to each step in a process in order to identify potential problems that may occur. A more human factors orientated version emerged in the form of the Human Error HAZOP, aimed at dealing with human error issues (Kirwan and Ainsworth 1992). In the development of another HEI tool Whalley (1988) also created a new set of guidewords, which are more applicable to human error. These Human Error guidewords are Not done, Repeated, Less than, More than, Sooner than, Later than, As well as, Mis-ordered, Other than and Part of. The guidewords are applied to each step in the HTA to determine any credible errors (i.e. those judged by the subject matter expert to be possible). Once the analyst has recorded a description of the error, the consequences, cause and recovery path of the error are also recorded. Finally, the analyst then records any design improvements to remedy the error. Human Error HAZOP appears to be a quick, easy to use and exhaustive technique. Similar to the SHERPA, its main weaknesses are that it is time consuming and also that some of the errors predicted using the tool could be questioned. Table 2 shows an extract of the Human Error HAZOP analysis performed in this study.

Task Step	Guideword	Error description	Consequence	Cause	Recovery path	Design improvements
3.2.2	Not Done	Pilot fails to enter new airspeed	Plane will not slow and may be travelling too fast for the approach	High workload Preoccupation with other landing tasks	3.2.1	Auditory prompt Warning - plane is travelling too fast
3.2.2	Less Than	Pilot does not turn the Speed/MACH knob enough	airspeed is not reduced enough and the plane may be travelling to fast	Poor control feedback Pilot inadequacy	3.2.1	Improved control feedback

Table 2 – Extract of Human Error HAZOP analysis

Task step	Error code	EEM	Description	PEM System cause	Consequence	Error reduction guidelines
3.2.2	PEP3	Action on wrong object	Pilot alters the airspeed using the wrong knob e.g. heading knob	Topographic misorientation Manual variability Mistakes alternatives Intrusion Similarity matching	The airspeed is not altered and the heading will change to the value entered	Ergonomic design of controls and displays Ergonomic procedures Training Clear labelling
3.2.2	PEP4	Wrong action	Pilot enters the wrong airspeed	Similarity matching Recognition failure Stereotype takeover Misperception Intrusion	Airspeed will change to the wrong airspeed	Training Ergonomic procedures with checking facilities Prompt system feedback

Table 3 – Extract of HEIST analysis

Human Error Identification in Systems Tool (HEIST)

HEIST (Kirwan 1994) is a technique that has similarities to a number of traditional HEI techniques such as SRK, SHERPA and HRMS (Kirwan 1994). The technique forms part of the HERA methodology (Kirwan 1998b). HIEST can be used by the analyst to identify external error modes via using the HEIST tables which contain various error prompt questions which are designed to prompt the analyst for potential errors. An example of a HEIST error identifier prompt would be, “Could the operator fail to carry out the act in time?” There are eight tables in total, under the headings of Activation/Detection, Observation/Data collection, Identification of system state, Interpretation, Evaluation, Goal selection/Task definition, Procedure selection and Procedure execution. The analyst applies each table to each task step from the HTA and determines whether any errors are credible or not. For each credible error, the analyst then records the system cause or psychological error mechanism and error reduction guidelines (both of which are provided in the HEIST tables) and also the error consequence. The methods main advantage is the use of error identifier questions which prompt the analyst for potential errors. However, the method does suffer from a number of domain transfer problems due to the fact that it was developed for the nuclear power industry. These weaknesses include the problem that the error identifier prompts are not all applicable to an aviation context. Examples of these include, “Does the signal occur at the appropriate time?” and “Will it be clear who must respond?” The error-reduction guidelines are also specific to the nuclear power industry. HEIST can also be time consuming to perform. Table 3 shows an extract of the HEIST analysis performed in this study.

Results

Each method was applied by the same analyst to a HTA of the landing task, “Land at New Orleans using the Autoland system’. To compute validity statistics, the predictions were compared with error data reported by pilots using the autoland system, which was supplied by Cranfield University. The signal detection paradigm was used as it has been found to provide a useful framework for testing the power of HEI techniques and has been used effectively for this purpose in the past (Stanton and Stevenage 2000). It is apparent that whilst this form of statistical analysis is unsophisticated, it is the most appropriate in terms of testing the accuracy of a HEI technique. The authors are also currently looking at other appropriate forms of statistical analysis, with Phi being one possible alternative. The signal detection paradigm uses the following four categories:

1) Hit – Predicted errors that actually have occurred
2) Miss – Failure to predict errors that have occurred
3) False Alarm – Predicted errors that have not occurred
4) Correct rejections – Correctly rejected errors that have not occurred

The signal detection paradigm can be used to calculate the sensitivity index (SI). This provides a value between 0 and 1, the closer that SI is to 1, the more accurate the techniques predictions are. Also calculated using the signal detection paradigm is the hit rate, which gives a rating in terms of how many hits V's misses the technique achieved, and the number false alarms which the technique predicted.

Analysis of the data revealed that SHERPA achieved the highest SI score, with 0.72. HEIST achieved a SI of 0.68 whilst the Human Error HAZOP achieved a SI of 0.61. These results imply that of the three methods used, SHERPA was the most accurate in terms of error prediction. The scores also indicate, however, that all three of the methods were moderately successful in their predictions. In terms of hit rate, which indicates how accurate the methods were in predicting 'real' errors. SHERPA again achieved the highest score, with an overall hit rate of 0.66, compared to HEIST's 0.54 and HAZOP's 0.43. This suggests that SHERPA was the most successful in predicting errors that had actually happened in the scenario under analysis and also made the least amount of misses (a miss being a failure to predict an error that has occurred). Finally, the total number of false alarms achieved by each technique revealed that SHERPA again performed the best of the three, with a false alarm total of only 27, compared to HAZOP's 42 false alarms and HEIST's 58 false alarms. Table 4 shows the results obtained from the analysis of each methods performance.

Method	**Sensitivity Index (SI)**	**Hit rate**	**False alarms**
SHERPA	**0.72**	0.66	27
HAZOP	**0.61**	0.43	42
HEIST	**0.68**	0.54	58

Table 4. SI values for SHERPA, HAZOP and HEIST

Conclusions

In conclusion, the results demonstrate that of the three HEI techniques, SHERPA performed the best when used to predict pilot error on a landing task using the autoland system. It can therefore be tentatively concluded that of the contemporary methods, SHERPA is the most suited for use in aviation. It appears that the SHERPA error taxonomy is the most suited to the tasks carried out by a civil aircraft pilot, with 'actions' and 'checks' being the most prominent tasks involved. It is also apparent that the different performance of the three methods is due to the constraints imposed on the possible errors that can be predicted by the methods. The possible errors that can be predicted by each method are determined by SHERPA's error taxonomy, Human Error HAZOP's guidewords and by HEIST's error identifier questions. For example, the guidewords used in the Human Error HAZOP method do not allow the analyst to predict an error such as, "Pilot enters airspeed using the heading knob instead of the speed/Mach knob", i.e. pilot presses wrong button. The SHERPA error taxonomy, however, prompts the analyst for this error, with 'A6 – Right action on wrong object'. The results indicate that existing HEI techniques, specifically developed for use in other domains (e.g. Nuclear Power), can be applied with some success in an aviation context. It appears that all of the three techniques, SHERPA, HEIST and Human Error HAZOP could be used effectively to predict error on modern day flight decks. This is quite encouraging, and suggests that with suitable development, existing techniques such as SHERPA can be used in the future to predict pilot error. This would of course only be viable if the three methods were to undergo substantial development. SHERPA for example would benefit from a more specific pilot behaviour based error taxonomy, whilst HEIST would need a whole new set of aviation specific error identifier questions to be developed. Similarly, the Human Error HAZOP method would benefit from a set of aviation task specific guidewords, such as 'Check wrong display' or 'Operate wrong control'. The goal for researchers now remains to investigate how these contemporary HEI methods can be improved and also the development and creation of new, aviation specific HEI methods. Human factors professionals face the challenge of determining exactly how these existing HEI methods can be developed so that they become efficient in predicting human error on modern day flight decks, and also the creation and validation of these new aviation specific methods.

Acknowledgements

The authors wish to acknowledge that this research was made possible through funding from the Department of Trade and Industry as part of the European EUREKA! Programme. Our thanks go to John Brumwell, Technical Manager of the CARAD Advanced Systems Programme at the DTI and Richard Harrison (now at QinetiQ) who have provided invaluable support to this research effort. Our thanks also go to Air2000, British Midland and JMC airlines in allowing us access to their pilots and to the pilots for taking the time to complete the questionnaire.

References

Annett, J.; Duncan, K. D.; Stammers, R. B & Gray, M. J. (1971) Task Analysis. Training Information No.6. HMSO: London.

BASE (1997) Boeing Airplane Safety Engineering, Statistical summary of Commercial Jet Airplane Accidents – Worldwide Operations, 1959-1996. In K. L McFadden & E. R Towell (eds.) *Aviation human factors: a framework for the new millennium*. Journal of Air Transport Management, 5, 177-184.

Embrey, D. E. (1986) SHERPA: A systematic human error reduction and prediction approach. Paper presented at the *International Meeting on Advances in Nuclear Power Systems*, Knoxville, Tennessee.

Kirwan, B. & Ainsworth, L. K. (1992) A Guide to Task Analysis, Taylor and Francis, London.

Kirwan, B. (1992a) Human Error Identification in Human Reliability Assessment. Part 1: Overview of approaches. Applied Ergonomics, 23 pp. 299-318.

Kirwan, B. (1994) A Guide to Practical Human Reliability Assessment, Taylor and Francis, London.

Kirwan, B. (1998b) Human Error Identification Techniques for Risk Assessment of High Risk Systems– Part 2: Towards a Framework Approach. *Applied Ergonomics*. 29 (5), 299-318.

Kletz, T. (1974) HAZOP and HAZAN: Notes on the identification and assessment of hazards. In C. D. Swann & M. L Preston (eds.) *Twenty five years of HAZOPs. Journal of loss prevention in the Process Industries*, vol 8 (6)

McFadden, K. L. (1993) An empirical investigation of the relationship between alcohol and drug related motor vehicle convictions and pilot flying performance. In K. L McFadden & E. R Towell (eds.) *Aviation human factors: a framework for the new millennium*. Journal of Air Transport Management, 5, 177-184.

McFadden, K. L. & Towell, E. R. (1999) Aviation human factors: a framework for the new millennium. Journal of Air Transport Management, 5, 177-184.

Nelson, W.R, Haney, L.N, Ostrom, L.T, Richards, R.E. (1998) Structured methods for identifying and correcting potential human errors in space operations, *Acta Astronautica*, vol 43, pp 211-222.

Stanton, N.A & Stevenage, S. V. (2000) Learning to predict human error: issues of acceptability, reliability and validity. In J. Annett & N. A. Stanton (eds.) *Task Analysis*. Taylor and Francis, London.

Swann, C. D. & Preston, M. L. (1995) Twenty five years of HAZOPs. Journal of loss prevention in the Process Industries, vol 8 (6), pp. 349-353.

Whalley (1988) Minimising the cause of human error. In B. Kirwan & L. K. Ainsworth (eds.) *A Guide to Task Analysis*. Taylor and Francis, London.