Irish Ergonomics Society

Irish Ergonomics Review

2015

Proceedings of the Irish Ergonomics Society Annual Conference 2015

Edited by

Chiara Leva

Leonard W. O'Sullivan

ISSN 1649-2102

Irish Ergonomics Review 2015 Proceedings of the Irish Ergonomics Society Annual Conference 2015

Published by The Irish Ergonomics Society

ISSN 1649-2102

TABLE OF CONTENTS

HUMAN FACTORS IN THE DIGITAL WORLD

S. Sharples

Faculty of Engineering, University Park, Nottingham, NG7 2RD, UK

Abstract

Digital technologies are changing the way we work, and the way we apply Ergonomics and Human Factors (E/HF). This talk will consider the theoretical and practical impact of novel technologies on the way in which people conduct tasks and jobs in contexts such as transport, healthcare and manufacturing.

The first trend considered is the increasing impact of automation in the workplace. E/HF specialist have studied automation for several decades, and traditional theories or models of automation consider 'levels' and identify the roles of the technology and the human in data acquisition and analysis, decision making and control. New technologies facilitate the move from automation to autonomy; it is therefore useful to consider how these changes in technology map on to our current theoretical approaches to automation, and how autonomous systems might be integrated with human roles in the future workplace.

A second trend is the impact that digital, personal technologies are having on the way that we work. The notion of the 'workplace' as a fixed desk with a personal computer is no longer the norm, and the proliferation of devices, as well as external societal and business influences are changing the way in which individuals complete their jobs. These changes are presenting opportunities and challenges and the key E/HF issues that emerge from these changes will be considered.

In addition to the impact that technologies are having on the way we work, technologies also offer potential to change the way we collect and analyse data as E/HF practitioners. New, less intrusive, sensor technologies offer the potential to monitor workplace performance in real time, and individuals are continually contributing to their 'contextual digital footprint'. This talk will present some examples of these new opportunities for collecting and analysing E/HF data, and consider the practical and ethical challenges that these changes in practice potentially present. Ultimately, the digital world is changing the way we work, and the way that we can apply E/HF theories and methods. By understanding these changes, and addressing some of the challenges to our knowledge and practice that these changes present, we can continue to develop novel E/HF theories, and provide leadership in the design of future workplaces.

PERCEPTIONS OF PERFORMANCE MANAGEPERCEPTIONS OF PERFORMANCE MANAGEMENT: A PILOT SURVEY OF AN AVIATION REGULATORY ENVIRONMENT

Seamus Clare and Dr Thomas Waldmann

Department of Design and Manufacturing Technology

University of Limerick, Ireland

Abstract

There can be many benefits associated with applying an appropriate strategy to guide an organisation into the future. A common theme amongst authors of strategic performance management material is often the degree of importance an organisation associates with its ability to measure its performance as informed by its organisational strategy. Indeed Kaplan and Norton (1992) who are credited with the development of a strategic performance management template (namely the balanced scorecard) compare a company's journey to the complex task of navigating and flying an aircraft.

Two of the main functions of the department featured in the case study are monitoring regulatory compliance of the domestic aircraft fleet and approved organisations and the support of a niche market that enables the operation of commercial aircraft at a number of global locations. By examining the attitudes towards and the understanding of performance management within a department which is not currently applying such a system, the research sets out to discover if a system such as a balanced scorecard were it to be implemented, could it have a positive impact in terms of effectiveness and efficiency on the services delivered? In addition to tighter corporate governance, "smarter working" requirements may eventually affect all organisations. It is also possible that harmonisation of goals and strategy could also contribute to more effective delivery of both regulatory and commercial services by the department.

Introduction

Bevan and Hood (2006), Modell (2009), Speklé and Verbeeten (2014) have reviewed a portion of the literature on performance management and measurement. Large scale data from aviation authorities and public sector organisations do not feature in the body of knowledge that results from alternative research Goddard (2010). The Chicago Convention ICAO (1944) places a legal obligation on each subscribing state to perform regulatory oversight of all registered aircraft that operate within state territories. Regulatory function and Government financial support may be provided in the same way as state funded General Medical Services, Legal, Immigration and Government departments that function on a "no-profit" basis. In contrast state aviation authorities may operate in part 'for profit' activities including air navigation and commercial aircraft leasing supports. Globally, aviation governance may be administered with supporting bureaucratic models and public sector departments. An ever increasing public awareness and more stringent governance has been a driver of change in the public sector Further research is needed on "the wider political ramifications of performance management outside of implementing organisations as it gets implicated in governance practices and regulation in various societal sectors" Modell (2012).

Background

The Aviation Authority (AA) in this pilot study is charged with regulation of safety standards (safety regulator function) and provision of air navigation services (service provider) in controlled airspace. The investigators are not aware of an effective method of accurate measurement or prediction of performance with the services department, service function when this pilot study was conducted. In addition to the State's aviation regulatory obligations, airworthiness services are provided to facilitate commercial aircraft operating at global locations as listed on the state civil register. A captive aviation market is apparent and in time may become vulnerable to change from environmental and market influences. This research presents perceived benefits of a "Balanced Scorecard" (BSC) system as a novel mechanism that may assist in the development of future strategic planning and delivery of regulation process of an aviation standards department.

The purpose of the research is to explore the perceived impact of implementing a multi-perspective performance management system in an AA department. The investigation aims to:

- Investigate understanding and expectation of performance management and measurement amongst the departmental staff.
- Investigate departmental staff's perception of a performance management system such as a BSC
- Identify potential obstacles and solutions in BSC implementation.

Performance Measurement and Performance Management

Wisniewski and Dickson (2001) seem to attribute the deficiencies in the performance management systems of old to the narrow focus of those previous architectures. Many of these frameworks were biased toward financial measures and automatically introduced an imbalance or a distortion of the overall view of how an entity was performing. According to Neely et al.

(1995) performance measurement is often discussed but it seems that it is rarely fully defined. Taticchi et al. (2010) sums up performance management succinctly with the Neely et al. (2002) definition, “A performance measurement and management system (PMS) is a balanced and dynamic system that enables support of a decision making process by gathering, elaborating and analysing information” This emphasises the evolutionary shift from a sole reliance in the past on financial information, to a combination of considering information from non-financial perspectives also.

In contrast Ittner and Larcker (1998) are critical of the balance proposition because “the BSC does not give a final score or end point result that managers can make decisions against and work towards”. Sundin et al. (2010) semantically examine the word “balance” in terms of grammatical context and discuss its possible interpretations in this context and connotation. ‘Dynamic’ according to Taticchi et al. (2010) refers to a system to monitor the status of goals and objectives which are affected by internal and external environmental factors. Valmohammadi and Servati (2011) consider a balanced scorecard as a means of offering a framework that supports the logical and consistent translation of a company’s strategy into workable outcomes. Many proven frameworks exist and can be tailored to support a performance management system. A theme amongst champions of performance measurement is the degree of importance associated with the ability to measure performance. According to Taticchi et al. (2010), “companies have understood that for competing in continuously changing environments, it is necessary to monitor and understand firm performances”. Kaplan and Norton (1992) analogy of an aircraft cock-pit compares the pursuit of successful business objectives to the “complex task of navigating and flying the aircraft”. As pilots need to be very much aware of the status of the aircraft parameters, managers similarly require performance information to support successful business objectives.

A combination of financial and non-financial perspectives is considered ‘first generation BSC’ Valmohammadi and Servati (2011). The BSC has evolved to address significant deficiencies in traditional management systems Kaplan and Norton (1996). The inability to link company long term strategies with short term action could be attributed to many management control systems. Kaplan and Norton (1996) confirm that this strategic dimension of the BSC was not a consideration for many companies when initially implementing the scorecard. Back then the motivation to implement a BSC appeared to be “to improve their performance management system” and “they achieved tangible but narrow results”. However Lawrie and Cobbold (2004) highlight the adverse effects of poor measure selection on the usefulness and adoption rates of balanced scorecards up to that point.

This second generation of the BSC according to Valmohammadi and Servati (2011) “emphasised the cause and effect relationship between measures and strategic objectives”. It would appear that the definition and application guidance material for a BSC was initially somewhat unclear. The brevity of the perspective statement outcomes, the birth of the objective, the development & recognition of causal linkage models and strategy maps were welcome features of this BSC (second) generation. Kaplan and Norton (1996). They also credit this innovation as “an improved measurement system to a core management system”

A later evolutionary offering of the BSC could be said to be a departure from the traditional four-perspective architecture. Many third generation scorecards feature what Lawrie and Cobbold (2004) term a "destination statement". Previously the initial design tasks for a BSC often involved development of objectives and measures. A destination statement could be best described as what an organisation will look like at a future agreed point in time. It would seem that by starting with the development of the destination statement, the task of measures & objectives selection and the promulgation and understanding of strategy is made easier with this starting point for BSC development. They also highlight their experience of public sector managers who have successfully applied a two-perspective scorecard simply using "activity" and "outcome" objectives related by causal linkages to meet the intent of their desired position in the future.

Has each company that implemented or attempted to implement a balanced scorecard fully realised their desired outcomes as originally designed? The degusted literature does not appear to yield a standard definitive understanding of a balanced scorecard. Organisations often have different interpretations of what a balanced scorecard is. Soderberg et al. (2011) also wonder if there is consistency amongst organisations that have implemented balanced scorecards despite the lack of unanimity amongst authors on the subject. Lawrie and Cobbold (2004) credit Kaplan and Norton (1992, 1993 and 1996) with the ability to motivate managers in many domains towards adopting and applying balanced scorecards, however they were "not seen as helpful with respect to operationalizing the BSC". Perhaps this degree of concern was the impetus for Soderberg et al. (2011) to develop a "five-level taxonomy" that could help support a better and more uniform understanding of BSC operationalization amongst practitioners. Additionally the idea of developing strategies in concert with performance management according to Modell (2012) should be considered against "the backdrop of political process unfolding in the external organisational environment".

Research Method

Identifying similar airworthiness standards' departments in other member state aviation authorities and gaining access to performance management structures and data, proved to be an inhibiting factor on the potential to conduct research on other similar organisation using the chosen research strategies. This work used an analysis of current literature in relation to the balanced scorecard and performance management. The main methodological approach used was a case study examining the elements below and comparing them where appropriate with the reviewed literature.

This strategy was further supported by a questionnaire survey tool distributed to a number of experienced staff that would be key stakeholders in the implementation of any strategic performance management initiative. The chosen research question and objectives relate to the areas of literature reviewed and the recurring findings and conclusions presented in these reviewed works. The department being studied represents a valuable opportunity to investigate the research problem further. Data from multiple sources from within the organisation and department was readily assembled through the use of a survey/questionnaire tool. Interviews (structured, semi-structured and unstructured) in addition to observational reporting was used to support and augment the primary strategy used. As required, archival and historical data (internal and external) was also examined. Applying supporting methods validated the data collected and made a contribution to a better understand and strengthened the research findings and conclusions.

Research activities conducted

Informal meetings held initially with senior managers and staff in the organisation and proved especially useful in establishing how the subject of strategic performance management was seen. The historical legacy issues that related to previous efforts to introduce performance management in the department and the reasons it was not successfully implemented were examined during some of the meetings. Some of the main points arising were, unclear objectives in the department, absence of structured performance management (PM), misinterpretation of what PM is, perceived limitations of PM, the importance of efficiency and effectiveness in the department and AA, appropriate metrics for the processes, understanding of a BSC, possible issues affecting PM implementation and suggested enabling activities. A questionnaire was used as the study's main instrument for data gathering. The literature reviewed and the points highlighted in the informal meetings collectively shaped the design of the distributed survey. The development of the questionnaire was based on an iterative process. The architecture of the questionnaire based tool comprises of six separate sections. The first stage involved piloting the questionnaire to four people. Two of the group were from the department, one from industry and one person from another department in the organisation. This activity resulted in the questionnaire being subdivided into six sections, some of the terminology was "laicized" and the order of the questions was changed to align with the research objectives. Following the embodiment of the changes resulting from the first iteration, the next revision of the document was piloted again with the same four individuals and was found to be satisfactory.

The questionnaire was distributed to thirty five colleagues in the department. Of this figure, thirty two responses were returned. The sample population was selected on the basis that they worked in the department for a period of greater than three years. Such criteria would ensure that the participants would have been exposed to all the department processes within that period. The questionnaires were distributed by hand and one month was the period allotted for their anonymous completion and return. The questionnaire applied a five-point 'Likert' rating scale to the possible answers. This tool was seen as a feasible option for data collection as it supports a quantitative approach using descriptive statistics for data interpretation. Open questions were also included to allow deeper feedback and allow respondents to fully convey their feedback and opinions in a manner which couldn't be achieved through closed-questions. The individual responses were coded and recorded on an excel spreadsheet. This format permitted each response to be shown against the individual questions. All statistical analysis was conducted using SPSS V20.

Findings

One of the objectives of the study was to investigate the understanding and expectation of performance management within department. The results are shown below;

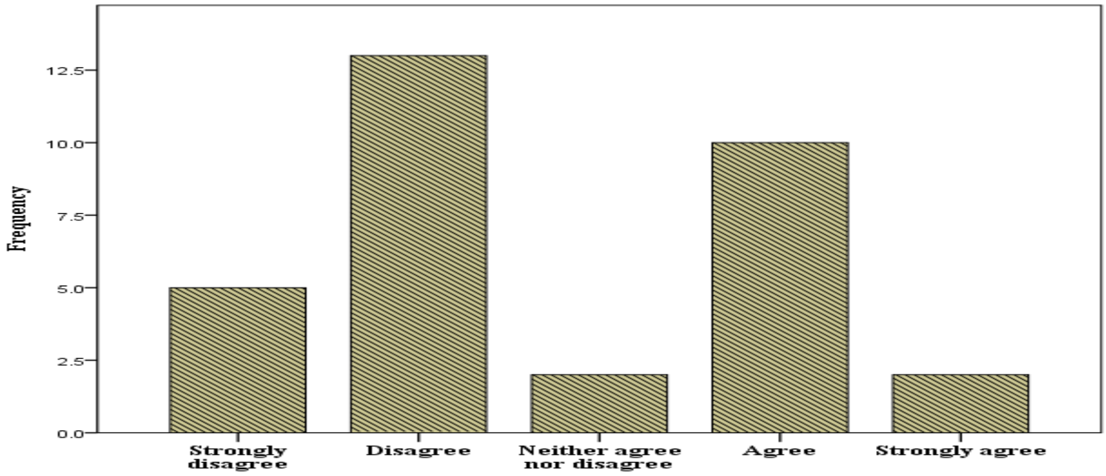


Figure 1. The objectives and goals pursued by department staff relating to the AA's strategy and vision are clearly defined in written form.

37.6% (n=12) agree or strongly agree that goals and objectives were clearly defined in written form. Over half of respondents (n=18) disagree with the statement. Of interest to note is the awareness of staff in relation to the expression of the department's goals and objectives.

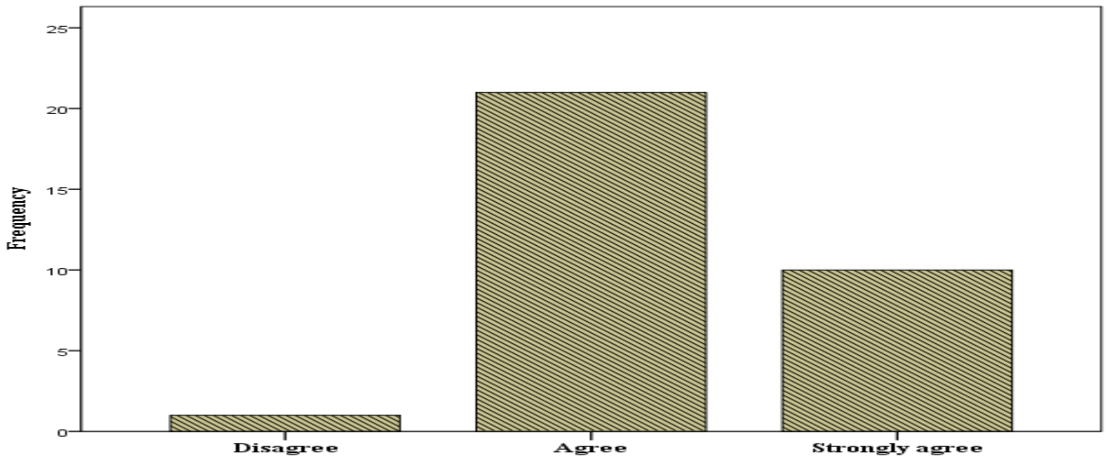


Figure 2.A structured system of performance management could support better decision making at all levels in the department/organisation

The data supports the benefits of structured performance management (PM). 65.6% (n=21) agree and 31.3% (n=10) strongly agree with the perceived ability of structured performance management (PM) to enhance the department's efficiency and effectiveness.

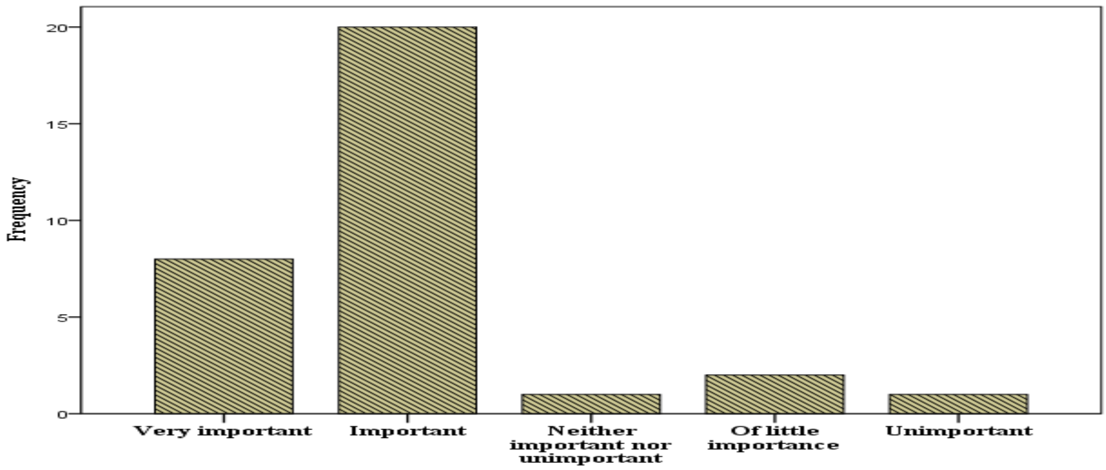


Figure 3.The department is regularly affected by organisational change. How important would you consider a system that provides up to date performance management information is to support such change?

The department is regularly affected by changes often precipitated by regulatory, technology-transfer and organisational evolutions. Respondents to the questionnaires placed great importance on up to date information when it comes to supporting organisational change, with 65% (n=20) considering it important and 25% (n=8) considering it very important.

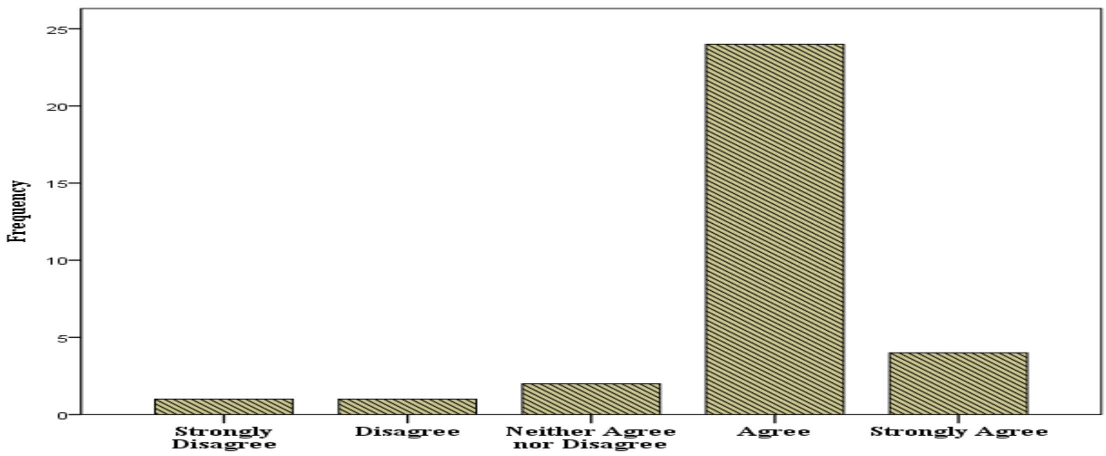


Figure 4.A performance management system utilizing a multi perspective template (such as a balanced scorecard) could support a clearer definition of objectives and their ultimate relationship with department/organisation strategy

The majority 75% (n=24) of staff reported that "A performance management system utilising a multi perspective template a (balanced scorecard) may support a clearer definition of objectives and their ultimate relationship with department/organisation strategy".

Table 1. Efficient internal processes are important elements that augment the delivery of effective services.

	Frequency	Valid Percent	Cumulative Percent
Agree	17	53.1	53.1
Strongly Agree	15	46.9	100.0
Total	32	100.0	

A critical component of any BSC is the 'internal perspective' of its composition. This sets out to measure and determine how well an organisation manages its operational processes. The architecture of the basic BSC can comprise of four perspectives or more (e.g. Customer, Innovation & Learning, Financial and Internal). The importance of efficient internal processes in augmenting the delivery of effective regulatory services is acknowledged by the combined unanimous response of 53.1% (n=17) in agreement and 46.9% (n=15) strongly agreeing.

Table 2. The idea of performance management where measures can be monitored and corrective action taken if necessary, could have a positive impact on the delivery of services.

	Frequency	Valid Percent	Cumulative Percent
Disagree	1	3.1	3.1
Neither Agree nor Disagree	5	15.6	18.8
Agree	17	53.1	71.9
Strongly Agree	9	28.1	100.0
Total	32	100.0	

The idea of performance management where measures can be monitored and corrective action taken if necessary is reflected in the result where 53.1% (n=17) agree and 28.1% (n=9) strongly agree with the statement. In contrast 3.1% (n=1) disagree and 15.6% (n=5) were undecided.

Table 3. A strategic performance management tool such as a balanced scorecard could easily identify "where we are now and where we are going in the future".

	Frequency	Valid Percent	Cumulative Percent
Strongly Disagree	1	3.1	3.1
Disagree	4	12.5	15.6
Neither Agree nor Disagree	4	12.5	28.1
Agree	19	59.4	87.5
Strongly Agree	4	12.5	100.0
Total	32	100.0	

Although the statement assumes the respondents have a degree of understanding of PM tools such as the BSC, the results show that almost three quarters of people (n=26) were in agreement with the statement posed in the survey.

The third and final aim featured in the paper considers the presence of possible impeding obstacles and mitigating solutions likely to support PM implementation.

Table 4. The department would benefit from the support of a cross-functional team with specific tasks, roles, responsibilities and processes to educate staff and promote the effectiveness of performance management.

	Frequency	Valid Percent	Cumulative Percent
Strongly Disagree	1	3.1	3.1
Disagree	2	6.3	9.4
Neither Agree nor Disagree	4	12.5	21.9
Agree	19	59.4	81.3
Strongly Agree	6	18.8	100.0
Total	32	100.0	

A dedicated cross-functional team tasked with educating staff and promoting the effectiveness of PM was endorsed by 59.4% (n=19) and 18.8% (n=6) who agreed and strongly agreed respectively. Coupled with the 12.5% (n=4) remained undecided, and 9.4% (n=2 & n=1) collectively disagreed with the statement.

Table 5. Any development of performance measures for the department should be linked to appropriate objectives and strategy.

	Frequency	Valid Percent	Cumulative Percent
Disagree	3	9.4	9.4
Neither Agree nor Disagree	2	6.3	15.6
Agree	18	56.3	71.9
Strongly Agree	9	28.1	100.0
Total	32	100.0	

A contention that future development of performance measures for the department should be aligned with appropriate objectives and strategy was evident in the results. This aspiration was not supported by 9.4% (n=3) of disagreeing respondents. Those in support were recorded as 56.5 (n=18) who agreed and 28.1% (n=9) who strongly agreed with linkage.

Table 6. In general, companies who engage in strategic planning do so in an effort to better manage the changing conditions which can disrupt the achievement of their long-range plans .

	Frequency	Valid Percent	Cumulative Percent
Strongly Disagree	1	3.1	3.1
Disagree	1	3.1	6.3
Neither Agree nor Disagree	7	21.9	28.1

	Frequency	Valid Percent	Cumulative Percent
Agree	15	46.9	75.0
Strongly Agree	8	25	100.0
Total	32	100.0	

46.9% (n=15) who agreed and a further 25% (n=8) strongly agreed that in general companies who engage in strategic planning do so in an effort to better manage the changing conditions which can disrupt the achievement of long range plans. This result is in contrast to 3.1% (n=1) disagreeing and 3.1% (n=1) who strongly disagreed. A total of 21.9% (n=7) neither agreed nor disagreed with the proclaimed merits of strategic planning.

Table 7. A performance management structure such as a balanced scorecard could improve communication and facilitate organisational learning in the department.

	Frequency	Valid Percent	Cumulative Percent
Strongly Disagree	1	3.1	3.1
Disagree	2	6.3	9.4
Neither Agree nor Disagree	2	6.3	15.6
Agree	23	71.9	87.5
Strongly Agree	4	12.5	100.0
Total	32	100.0	

Communication and organisational learning are synonymous with progressive organisations. The data illustrates that a majority of 71.9% (n=23) and 12.5% (n=4) of participants (agreed and strongly agreed respectively) supported the notion that a performance management (PM) structure such as a balanced scorecard (BSC) could improve communication and facilitate organisational learning in the department.

Discussion

It was evident from the data in this study that the responses supported the development of PM and any future development of measures should be linked to an appropriate organisational strategy. It could be said that Kaplan and Norton (1993) generalisation that many measures "operational and physical" still exist for "local activities" in organisations. The airworthiness authority (AA) may be no different. Kaplan and Norton (1993) highlight the ability of the scorecard in terms of its measures being grounded in an organisation's strategic objectives. The findings of this research indicate that a large proportion of the department's staff are in agreement in principle with the under-lying ethos of a balanced scorecard and recognise the reliance on its strategy/objective/goal inter-relationships.

The research has shown that an AA applies complex metrics such as levels of compliance, number of incidents, number of accidents and results from European and International audit programmes to support their quantification of effectiveness The findings of this pilot study are in keeping with previously reported research; Kaplan (1992), Kaplan (1993), Kaplan (1996), Kaplan (2001), Sundin (2010), Valmohammadi (2011), Wisniewski (2001). The survey findings have highlighted the awareness of the perceived importance of efficiency and effectiveness amongst the respondents (n=32) who indicated that the resources that support the services offered by the department are not always applied to best effect in their opinion.

The vast majority (n=27) of the questionnaire respondents felt that a performance management structure such as a balanced scorecard may improve communication and facilitate organisational learning within the department. As a conduit for communication the balanced scorecard has the ability to link long term strategic objectives with short term actions.

Conclusion

The investigation sets out to determine; the understanding and expectations of performance management (PM) amongst department staff, the surveyed perception of a PM system such as a BSC and obstacles to PM implementation (n=32).

The results indicate the organisation's strategy was considered to be well documented by those who participated in the study (37.6%, n=12, collectively agree or strongly agree). The survey results also showed that a structured PM system could contribute to better decision making within the organisation (31.3%, (n=10) strongly agree and 65.6%, (n=21) agree). Additionally a large proportion of staff (75%, n=24) were in agreement with the idea that a PM system such as a balanced scorecard (BSC) could support improved definition of objectives and their relationship with the organisation's strategy. Many of the perceived benefits of performance management are understood in the department and are further reinforced with the finding that 81.9% (n=23) of those surveyed felt that a balanced scorecard could easily offer strategic direction for the department/organisation.

Therefore it would appear;

- There is already a strong understanding of the concept of performance management within the department featured in the case study.
- The data suggests a positive perception of PM amongst respondents.

Finally the evidence in the data set concludes the department may ultimately benefit from an information initiative prior to implementing a performance management (PM) system. Promulgation of information within an organisation is extremely important. Preconceptions can support such constraints as poor attitude and lack of motivation which can become implementation obstacles, as considered by Garvin (1993) in Harvard Business (1998). As a result, educating staff prior to implementing initiatives for PM is considered a key action in support of a successful introduction. This was endorsed by 59.4% (n=19) and 18.8% (n=6) who agreed and strongly agreed respectively.

In conclusion the case study data supports the contention that there is a reasonable understanding of the concept of performance management amongst the group employed in the featured department of the AA. The benefits of performance management and measurement and the balanced scorecard strategic management as defined by Neely et al. (1995), Harvard Business (1998) and Kaplan and Norton (199), (1993) and (1996) are already appreciated by the majority of staff in this department. The findings of this pilot support the implementation of a multi-perspective performance management system amongst many of those who participated in the study.

Disclaimer

The authors' views expressed in this publication do not necessarily reflect the views of the Aviation Authority featured as the subject of the case study or the European Aviation Safety Agency.

References

Bevan, G. and Hood, C. (2006) 'WHAT'S MEASURED IS WHAT MATTERS: TARGETS AND GAMING IN THE ENGLISH PUBLIC HEALTH CARE SYSTEM', Public Administration, 84(3), 517-538.

Goddard, A. (2010) 'Contemporary public sector accounting research – An international comparison of journal papers', The British Accounting Review, 42(2), 75-87.

Harvard Business, S. (1998) Harvard Business Review on knowledge management, Boston, Mass: Harvard Business School.

Ittner, C. D. and Larcker, D. F. (1998) 'Innovations in Performance Measurement: Trends and Research Implications', Journal of Management Accounting Research, 10, 205-238.

Kaplan, R. S. and Norton, D. P. (1992) 'The Balanced Scorecard--Measures That Drive Performance', Harvard Business Review, 70(1), 71-79.

Kaplan, R. S. and Norton, D. P. (1993) 'Putting the Balanced Scorecard to Work', Harvard Business Review, 71(5), 134-147.

Kaplan, R. S. and Norton, D. P. (1996) 'Using the Balanced Scorecard as a Strategic Management System', Harvard Business Review, 74(1), 75-85.

Lawrie, G. and Cobbold, I. (2004) 'Third-generation balanced scorecard: evolution of an effective strategic control tool', International Journal of Productivity and Performance Management, 53(7), 611-623.

Modell, S. (2009) 'INSTITUTIONAL RESEARCH ON PERFORMANCE MEASUREMENT AND MANAGEMENT IN THE PUBLIC SECTOR ACCOUNTING LITERATURE: A REVIEW AND ASSESSMENT', Financial Accountability & Management, 25(3), 277-303.

Modell, S. (2012) 'The politics of the balanced scorecard', Journal of Accounting & Organizational Change, 8(4), 475-489.

Neely, A., Gregory, M. and Platts, K. (1995) 'Performance measurement system design', International Journal of Operations & Production Management, 15(4), 80-116.

Soderberg, M., Kalagnanam, S., Sheehan, N. T. and Vaidyanathan, G. (2011) 'When is a balanced scorecard a balanced scorecard?', International Journal of Productivity and Performance Management, 60(7), 688-708.

Speklé, R. F. and Verbeeten, F. H. M. (2014) 'The use of performance measurement systems in the public sector: Effects on performance', Management Accounting Research, 25(2), 131-146.

Stamatis, D. H. (2003) Guidelines for failure mode and effects analysis for automotive, aerospace and general manufacturing industries, Richmond Hill, Ont; London: Dyadem Press.

Sundin, H., Granlund, M. and Brown, D. A. (2010) 'Balancing Multiple Competing Objectives with a Balanced Scorecard', European Accounting Review, 19(2), 203-246.

Taticchi, P., Tonelli, F. and Cagnazzo, L. (2010) 'Performance measurement and management: a literature review and a research agenda', Measuring Business Excellence, 14(1), 4-18.

ter Bogt, H., Budding, T., Groot, T. and van Helden, J. (2010) 'CURRENT NPM RESEARCH: DIGGING DEEPER AND LOOKING FURTHER', Financial Accountability & Management, 26(3), 241-245.

Valmohammadi, C. and Servati, A. (2011) 'Performance measurement system implementation using Balanced Scorecard and statistical methods', International Journal of Productivity and Performance Management, 60(5), 493-511.

Wisniewski, M. and Dickson, A. (2001) 'Measuring Performance in Dumfries and Galloway Constabulary with the Balanced Scorecard', The Journal of the Operational Research Society, 52(10), 1057-1066.

STAKEHOLDER INVOLVEMENT IN EVALUATION: LESSONS LEARNED IN THE A-PIMOD PROJECT

J. Cahill and T.C. Callari

CHIS, School of Psychology, Trinity College Dublin, Dublin, Ireland

Abstract

The objective of this paper is to demonstrate the importance of utilizing a stakeholder evaluation approach in Ergonomics/ Human Machine Interaction (HMI) research. Specifically, this paper will introduce the stakeholder evaluation approach advanced in the A-PiMod project and the lessons learned in relation to the application of this approach. Further, it will provide recommendations for practitioners in relation to the adoption of this approach in Ergonomics research.

Introduction

Introduction to the paper

The objective of this paper is to demonstrate the importance of utilizing a stakeholder involvement approach to evaluation, in Human Factors/Ergonomics HF/E) and Human Machine Interaction (HMI) research. Overall, the paper focuses on the requirement for HF/E/HMI practitioners to properly collaborate with end users and project members, as conceptualised in the A-PiMod 'Community of Practice' (COP).

First, a background to automation systems and the associated HF problems is provided. The methodological background to this research is then presented. Following this, the A-PiMod project is introduced. The stakeholder approach to evaluation (and allied A-PiMod COP concept) advanced in this project is outlined. This is followed by a description of the specific collaborative evaluation techniques and HMI design/evaluation methods applied in the project, and the provisional results of this research. The benefits and challenges of this approach are then discussed. Recommendations for practitioners are then provided. Lastly, some conclusions are drawn.

Introduction to Research Problem

Crew information needs (and the requirement for workload and decision support) vary according to the crew composition, the specific experiences of the two crew members (i.e.

familiarity with type, familiarity with route and time elapsed since last flown that route), and the specific flight circumstances on the day (traffic, weather etc). With increasing flight hours, fatigue and forecast increased traffic growth, all crews can benefit from an "experience aid".

Given automation advances over the last decade, Pilots share responsibility for different flight tasks with cockpit systems. Adaptable systems are systems which require human delegation of task and 'function authority' to automation during real-time operational performance (Kaber & Prinzel, 2006). Adaptive automation (AA) is defined as a 'form of automation that allows for dynamic changes in control function allocations between a machine and human operator based on states of the collective human–machine system (Hilburn, Byrne, & Parasuraman, 1997; Kaber & Riley, 1999).

The air accident and flight safety literature reports on the many still-open issues in relation to automation design. For example: Flight Air France 447 (2009), Flight Spanair 5022 (2008), Flight Helios Airways HCY 522 (2005), Flight China Airlines 140 (1994), and Flight Air Inter 148 (1992). Critically, several human factors problems have been documented. This includes: automation surprises, workload concerns and issues pertaining to over-reliance on automation.

Evaluation Approaches/Methods

HCI/HMI methods

The HMI literature defines a range of formal and informal methods for the design of human friendly technology adopting a 'User-Centered Design' methodology (Cooper, 2007; Preece, Rogers, & Sharp, 2007; Constantine & Lockwood, 1999; Hackos & Reddish, 1998). Typically, formal HMI methods start with an analysis of the existing task (Preece, et al., 2007). Generally, this takes the form of structured or semi structured interviews with end users. This research informs the specification of user requirements. Following from this, prototypes are designed - mostly without the participation of end users. Users are then involved in the evaluation of prototypes. In the aviation context, this typically involves some form of simulator evaluation.

Formal HMI methods have been the subject of much debate in the HCI literature. Specific challenges have come from the fields of Ethnography and Participatory Design. Ethnographers argue that classical HCI methods fail to address the social aspects of work (Hutchins 1995; Vicente 1999). Participatory design theorists have questioned the separation between design and evaluation in formal methods (Bødker & Buur, 2002). Central to Participatory Design theory is the idea that Usability Engineers design 'with' end users, as opposed to 'for' them. Accordingly, users are active participants in the design process (Bannon & Bødker, 1991, Bødker & Grønbæk, 1996).

Stakeholder Involvement approaches to Evaluation/Community of Practice Approaches

Stakeholder involvement in programme evaluation has been recognised as one of the most effective approaches to enhancing the use of evaluation findings, and ensuring the validity of the evaluation activities (Brandon & Fukunaga, 2014). Stakeholder involvement is defined as the participation of (programme) stakeholders in any phase of an evaluation (O'Sullivan, 2012). Stakeholder involvement can vary with regard to diversity in stakeholder selection for participation, the control of technical evaluation decisions and the depth of stakeholder participation in the programme/project evaluation process (Cousins, Whitmore, & Shulha, 2013). Stakeholders are conceived as invaluable source of knowledge, perspectives, information on context and needs. Drawbacks of stakeholder involvement are also reported. This includes the feasibility of implementing a successful participative study. For example, time, cost, involvement from (disadvantaged) groups and skills required from an evaluator in facilitation and "good listening" (Fitzpatrick, Sanders, & Worthen, 2010).

The involvement of stakeholders to accomplish given tasks by participating in common activities has been central to 'Community of Practice' concepts (Wenger 1991, 1998). 'Community of Practice' members engage in a set of relationships over time around some particular area of technical knowledge or skill associated with the given tasks. This allows the members of a specific 'Community of Practice' to generate a sense of joint enterprise and identity by sharing a practice - doing things together, developing a sense of place, common goals. In Wenger's analysis, three characteristics are crucial to define a 'Community of Practice': (1) the '*domain*' – which specifies the identity of COPs with the specific competence and commitment the stakeholders engage; (2) the '*community*' – stakeholders build their relationship interacting in joint activities, sharing information and common objectives and learning from each other; and (3) the '*practice*' – stakeholders share a repertoire of resources (experiences, stories, tools, ways of addressing recurring problems) which help forming the practice with time and sustained interaction (Wenger, et al., 2002).

Stakeholder Evaluation in A-PiMod

Introduction to A-PiMod Project

The Applying Pilots' Model for Safer Aircraft (A-PiMod) project aims to address problems relating to crew/automation teamwork and workload management. The high level goal of the project is to design a new adaptive automation concept based on a hybrid of three elements – (1) Multi-Modal Pilot Interaction, (2) Operator Modeling, and (3) Real-Time Risk Assessment. The specific objective of TCD's research is to validate the A-PiMod concepts and technologies from a functional and operational/safety perspective. This spans requirements specification/validation, prototype design and evaluation, and the final evaluation of safety/operational impact.

Overall HMI Design/Evaluation Approach

The high level HMI design/evaluation methodology adopted combines formal HMI design/evaluation activities (i.e. interviews and simulator evaluation), informal HMI design/evaluation approaches (i.e. participatory design activities: co-design/evaluation of prototypes), along with an integrated stakeholder approach to evaluation (linked to the Community of Practice concept). Currently, thirteen COP sessions have been completed, while a further two are in progress. These are summarised in Appendix 1. In addition, the first round of simulator evaluation has been undertaken.

The A-PiMod Community of Practice Concept

The concept of COP proposed by Wenger underpins the A-PiMod 'Community of Practice' approach (this is illustrated in Appendix 2). The A-PiMod 'Community of Practice' involves stakeholders who share technical knowledge and skills associated with relevant functions in the Air Traffic Management (ATM system), and broader aviation related domain. Overall the role of participants in the A-PiMod COP concept is characterised as a 'participatory' approach. Members engage in a range of validation/evaluation activities on a continuous/regular basis, through the run-time of the project.

Stakeholder & Researcher/Evaluator Roles

The panel of stakeholders in A-PiMod include both 'primary users' (i.e. internal stakeholders representative of each project partner) and 'all legitimate groups' (i.e. external stakeholders representative of the aviation-related industry and Flight operational system). The current COP panel comprises 15 participants (i.e. ten internal stakeholders and five external stakeholders). It is anticipated that TCD may recruit additional external stakeholders, for the purpose of specific validation exercises.

The role of internal stakeholders involves:

- Assessing what is technically feasible in relation to user requirements
- Providing input based on their own domain knowledge
- Providing feedback to their company regarding user needs, user requirements and validation findings
- Facilitating the handover of requirements and prototypes

The role of external stakeholders involves:

- Supporting the specification and evaluation of A-PiMod system concepts and technologies
- Ensuring A-PiMod concepts and technologies address real world operational and safety requirements
- Providing comments and information from direct experience and practice which can be taken into account in terms of the design of the A-PiMod system
- Evaluating the potential safety/operational impact

TCD's role is that of a facilitator, to promote a supportive learning environment in which the participants share their expertise and learn from the group collaboration.

To support the recruitment process, TCD prepared a background document to familiarize the COP stakeholders with the project and, in particular, to set expectations concerning the role of the A-PiMod 'Community of Practice', and associated evaluation activities. In addition, a Profile Form was circulated to all stakeholders to capture basic personal details and information concerning their professional expertise. This allowed evaluators/researchers to map the range of stakeholder competency/expertise in the ATM domain - linking to stakeholder evaluation goals/requirements. The Radar Diagram below (see Figure 1) shows the two overlaying levels of expertise both from the internal and external stakeholders. The composition of the internal stakeholders is represented in blue, while the composition of the external stakeholder is represented in amaranth. The red dotted line corresponds to the 2-level expertise.

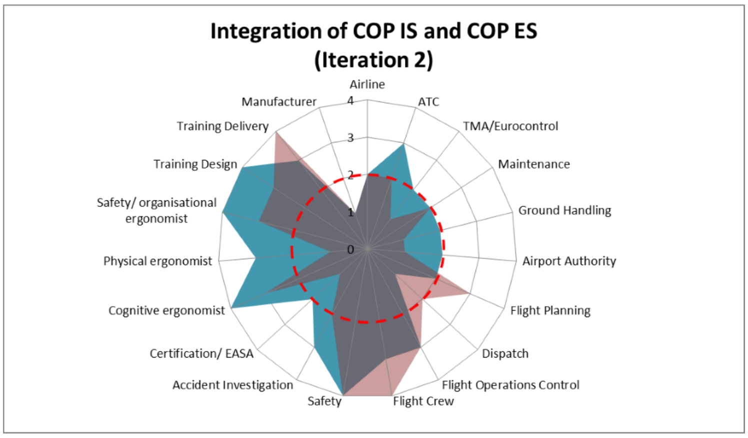


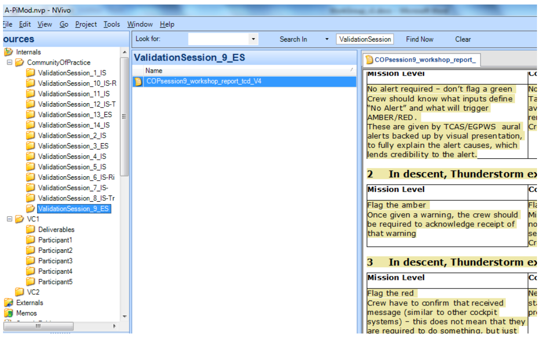
Figure 1. Current state of stakeholder competency knowledge in A-PiMod

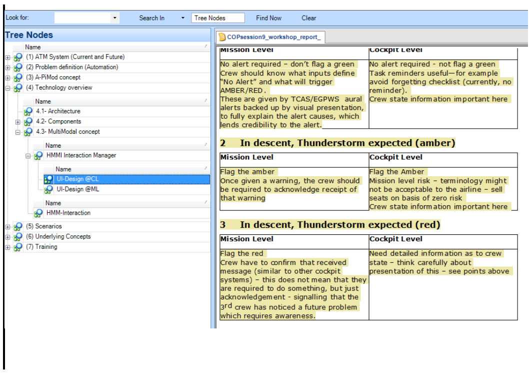
Stakeholder involvement in the evaluation process

Stakeholder participation involves consultative interaction along with engagement in technical research tasks (Cousins, et al, 2012). Stakeholders provide feedback both formally (involvement in specific COP design/evaluation sessions) and informally (ongoing email feedback, documentation of issues etc). On completion of each COP validation exercise, TCD have circulated reports summarising the topics/themes addressed, the methods used, the primary findings, the agreement reached by the participants, and the proposed action items. Participants involved in each research session are asked to review the contents of each report. Following the assessment and integration of participant feedback, each report is signed off. Results from all validation sessions are shared in a designated area of the project SharePoint (i.e. electronic repository) and signed off by the members of the A-PiMod 'Community of Practice'. Thus members influence the scope of this adaptive automation research, and the research outcome.

Research Methods

Different techniques/sources of evidence (among which semi-structured interviews, participatory workshops {some of which involved co-design activities and the completion of questionnaires}) have been used. These are depicted in Appendix 3. Overall, these exercises support (1) the definition of the adaptive automation concept and the associated A-PiMod architecture, (2) the specification of human factors/user requirements for the new system, (3) the specification of specific user interface design requirements for the mission level and cockpit level displays, and (4) the elaboration of potential A-PiMod demonstration scenarios. Further, they serve as an input to planning activities in relation to formal HMI evaluation activities (i.e. Simulator Evaluation Sessions 1 and 2). The fifteen COP validation exercises reports/findings have been categorised and analysed in the proposed categories of the A-PiMod coding frame (See Appendix 4 for details). Data recording and analysis is being undertaken with the assistance of a Computer-Assisted Qualitative Data Analysis Software (CAQDAS) tool - NVivo (© QSR International, V.8) (Bazeley, 2007). Figures 2 and 3 below show the current coding activity, categorisation and analysis performed in the NVivo A-PiMod project.





Figures 2 and 3. Data recording and analysis (coding activity) with NVivo

Preliminary Results

Several categories of human factors problems have been identified (see Appendix 5). The goal is to support crew in situations when they may need help/back-up – irrespective of experience, and/or in situations when the crew has less experience, and/or in situations where the crew is fatigued or under pressure. Overall, this links to an assessment of crew state. The assessment of crew state is not just about workload/situation awareness - it's about the crew experience, flight hours, familiarity with route, when last flown there, training background etc. Further, it takes into account real-time crew behaviour (i.e. monitoring of crew interaction with systems – touch, voice and traditional controls, gaze data and so forth). The crew and automation co-operate in relation to mission level decisions. Automation can be conceptualised as a virtual team-member. A-PiMod continuously monitors the operational situation and the allied crew/automation state, to determine the best distribution of task activity between the crew and automation. A-PiMod provides the best support to crew in terms of flagging potential risks, and providing guidance in relation to managing those risks (i.e. course of action/flight plan). If the crew state is not optimum, both crew members should be paying more attention, looking for help (i.e. using the auto-pilot), tracking each other's activities and level of situation awareness. During the flight, there may be different levels of crew state monitoring. For example, (1) passive support, (2) active support and (3) intervention/over-ride.

Discussion

The participation of stakeholders in the Community of Practice provides a strong link to the real world, in terms of understanding (1) automation issues, and (2) the capacity of technology to address these issues. Also, this provides an effective mechanism to enable project co-ordination and teamwork. The involvement of internal stakeholders has helped forge a link between the human factors/user research and the technical research. The importance of involving external stakeholders (i.e. pilots) cannot be understated. This involvement has been critical to the collection of user requirements and the emerging definition of the A-PiMod concept. The implementation of Community of Practice research is not straightforward. This requires the advancement of a 'working relationship' with community members (i.e. trust and teamwork), the set-up and acceptance of communication/information sharing practices and the establishment of a decision making process. All of this takes time. Further, the adoption of a participatory approach can make decision making slow. Researchers must balance the needs of individual internal stakeholders (who may focus on their company's own product development research agenda), and external stakeholders (who represent the operational need).

Recommendations for Practitioners

Overall, the ongoing involvement of stakeholders in A-PiMod evaluation activities as part of the COP, as yielded several lessons learned. These have been formulated has recommendation for practioners, and are detailed in Table 1 below.

T**able 1.** Recommendations for Practioners.

#	Category	Recommendations
1	**Set-up of 'Community of Practice'**	Benefits from formal set up (i.e. Treated as areal community). Publish list of names/emails. Preparation of an 'information package' to explain project goals, stakeholder role, set expectations re involvement. The available resources for stakeholder involvement should be clearly stated.
2	**Profile of members**	In relation to both internal and external stakeholders, it is important to have a spectrum of roles/profiles. Stakeholder recruitment is an on-going process- it responds to the specific need of the project development-the different profiles/expertise required

		may vary over time.
3	**Capture Profile Data**	It is useful to obtain profile information in advance(1) to assess the competency of the group as a whole and (2) to help weight feedback
4	**Set Expectations in terms of Level of Participation**	Setting realistic expectations in terms of level of participation and control impacts on behaviour and level of interest. Roles and responsibilities can evolve over time – participants should be made aware of this.
5	**Role of Researcher/Evaluato r**	The researcher/evaluators are in charge of creating an ongoing engagement between stakeholders and evaluators. We recommend adopting a 'brokerage role' between internal/external stakeholders. This is underpinned by quality communication and the establishment of good working relationships between the Researcher and internal/external stakeholders (i.e. trust and teamwork).
6	**Communication techniques/Eliciting feedback**	Active and on-going communication between evaluators/researchers and stakeholders leads to stronger evaluation designs, enhanced data collection and analysis, and results that stakeholder scan understand and use. Participation should include consultative interaction along with engagement in technical research tasks. Use of formal and informal communication mechanisms. Set-up of mailing lists– easy communication with members. Set up of information repository to store data/documents– and facilitate document sharing.
7	**Planning Research Sessions**	Each evaluation session should have a clear research objective, method and agenda/brief. It is useful to elicit feedback in advance of the session (i.e. set tasks to each stakeholder), and to publish the findings of this in advance of the session.

8	**Weighting feedback**	Set expectations about how feedback is weighted and used. Establish the consensus view. Publish a final report for each session, which includes feedback from all partners (i.e. transparency in decision making).
9	**Documenting and analysing research**	Each evaluation session should be formally reported, and stakeholders should provide feedback (and signoff on agreement/level of consensus). An accurate and transparent picture of data recording should be provided, along with an audit of the data analysis process as a whole.NVivo (Computer-Assisted Qualitative Data Analysis Software {CAQDAS) is useful for this purpose.
10	**Learning Environment to Share Ideas**	The creation of an inclusive learning environment where members share ideas necessitates an appropriate setting (and potentially technology). In A-PiMod this has been mostly remotely telephone/web mediated (i.e. with WebEx), although some person to person interviews have been undertaken. Overall, person-to-person interaction has proved the most fruitful.

Conclusions

Overall, the stakeholder evaluation/validation approach adopted has facilitated the preliminary specification and evaluation of a new adaptive automation concept. Specifically, the integration of a range of formal and informal HMI methods, with a stakeholder evaluation approach has proved effective in terms of enabling both operational and safety validation. The participation of stakeholders in the Community of Practice provides a strong link to the real world –in relation to (1) understanding automation issues, and (2) the capacity of technology to address these issues. Critically, the emerging adaptive automation concept is predicated on feedback in relation to flight crew experience with automation (and associated problems).

Acknowledgements

The research leading to these results/preliminary outcomes has received funding from the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement N. 605141 - Applying Pilot Models for Safety Aircraft (APiMod) Project.

We would like to thank member of the A-PiMod Project Team and our COP members.

References

Bannon, L., & Bødker, S. (1991). Beyond the Interface: Encountering Artefacts in Use. In J.M. Carroll (Eds.), *Designing Interaction: Psychology at the Human-Computer Interface* (pp. 227-253). New York: Cambridge University Press.

Bazeley, P. (2007). *Qualitative Data Analysis with NVivo*. London: SAGE Publications.

Bødker, S., & Burr, J. (2002). The Design Collaboratorium. A Place for Usability Design. *ACM Transactions on Computer Human Interaction, 9*(2), 152-169.

Bødker, S., & Grønbæk, K. (1996). Users and Designers in Mutual Activity. An analysis of cooperative activities in systems design. In Y. Engeström and D. Middleton (Eds.), *Cognition and Communication at Work* (pp. 130-158). Cambridge: Cambridge University Press.

Brandon, P. R., & Fukunaga, L. L. (2014). The State of the Empirical Research Literature on Stakeholder Involvement in Program Evaluation. *American Journal of Evaluation, 35*(1), 26-44.

Cahill, J., & Callari, T. C. (2015). *A Novel Human Machine Interaction (HMI) Design/Evaluation Approach Supporting the Advancement of Improved Automation Concepts to Enhance Flight Safety*. In D. de Waard, J. Sauer, S. Röttger, A. Kluge, D. Manzey, C. Weikert, A. Toffetti, R. Wiczorek, K. Brookhuis, and H. Hoonhout (Eds.). Proceeding of the Ergonomics Society Europe Chapter 2014 Annual Conference "Human Factors in high reliability industries" Lisbon, Portugal. Available from http://www.hfes-europe.org/human-factors-high-reliability-industries-2/

Constantine, L.L., & Lockwood, L.A.D. (1999). *Software for Use: A Practical Guide to the Models and Methods of Usage-centered Design*. MA: Addison-Wesley.

Cooper, A. (2007). *The inmates are running the asylum* (7th ed.). Indianapolis: SAMS Publishing.

Cousins, J. B., Whitmore, E., & Shulha, L. (2013). Arguments for a Common Set of Principles for Collaborative Inquiry in Evaluation. *American Journal of Evaluation, 34*(1), 7-22.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2012). *Program evaluation: alternative approaches and practical guidelines* (4th ed.). Upper Saddle River, US: Pearson Education..

Flight Air Inter 148 (1992, January 20). *Final Report on the accident on 20th January 1992 involving an Airbus A320 registration F-GGED operated by Air Inter Airlines in Vosges*

Mountains (near Mont Sainte-Odile). Retrieved from Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA). http://www.bea.aero/docspa/1992/f-ed920120/htm/f-ed920120.html

Flight China Airlines 140 (1994, April 26). *Final Report on the accident on 26th April 1994 involving an Airbus Industrie A300B4-662R registration B1816 operated by China Airlines at Nagoya Airport (Published July 19, 1996)*. Retrieved from Aircraft Accident Investigation Commission. http://www.skybrary.aero/bookshelf/books/808.pdf

Flight Helios Airways HCY522 (2005, August 14). *Final Report on the accident on 14thAugust 2005 involving a Boeing 737-31S registration 5B-DBY operated by Helios Airways at Grammatiko, Hellas (Published November 2006)*. Retrieved from Air Accident Investigation & Aviation Safety Board (AAIASB). http://www.moi.gov.cy/moi/pio/pio.nsf/All/F15FBD7320037284C2257204002B6243/$file/FINAL%20REPORT%205B-DBY.pdf

Flight Spainair 5022 (2008, August 20). *Final Report on the accident on 20th August 2008 involving a McDonnell Douglas DC-9-82 (MD-82) registration EC-HFP operated by Spainair at Madrid-Barajas Airport (Published October 8, 2008)*. Retrieved from Comisión Investigatión de Accidentes e Incidentes de Aviación Civil (CIAIAC). http://www.fomento.es/NR/rdonlyres/EC47A855-B098-409E-B4C8-9A6DD0D0969F/107087/2008_032_A_ENG.pdf

Flight AF 447 (2009, June 1). *Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro (Published July 2012)*. Retrieved from Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA). http://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf

Hackos, J.A., & Redish, J.C. (1998). User and Task Analysis for Interface Design. New York: Wiley Computer Publishing, John Wiley & Sons Inc.

Hilburn, Byrne, & Parasuraman, 1997 Horwood Limited (publishers), John Wiley & Sons (Halsted Press)

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Kaber, D. B., & Riley, J. M. (1999). Adaptive automation of a dynamic control task based on secondary task workload measurement. International Journal of Cognitive Ergonomics, 3, 169–187

Kaber, D.B., & Prinzel, L.J. (2006). Adaptive and Adaptable Automation Design: A Critical Review of the Literature and Recommendations for Future Research. NASA/TM-2006-214504

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*: Cambridge University Press.

O'Sullivan, R. G. (2012). Collaborative Evaluation within a framework of stakeholder-oriented evaluation approaches. *Evaluation and Program Planning, 35*, 518-522.

Preece, J., Rogers, Y., & Sharp, H. (2007). Interaction Design: Beyond Human-Computer Interaction (2nd ed.). West Sussex, UK: John Wiley & Sons Ltd.

Vicente, K.J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge university press.

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Boston: Harvard Business Press.

Appendices

Appendix 1: Stakeholder Evaluation Research

Table 2. COP Schedule of Activities: Year 1

Schedule	**Topic**	**Who Involved**	**Status**
March 2014	Automation concept, architecture and components description and use scenario	All project partners (not specifically COP)	Complete
March 12 2014	Automation concept, architecture and components description and use scenario	COP/IS (SYM, DLR, TCD)	Complete
May 11/12 2014	Automation problems and eliciting improved automation concept	COP/ES	Complete
July 15 2014	VETO question (crew/automation	COP/IS - ALL	Complete

	roles and responsibility)		
June 18 2014	Crew model, workload concepts and situation awareness/assessment concept	COP/IS (DLR, OFF, SYM and TCD)	Complete
July 9 2014	Risk Assessment concept and tools	COP/IS (TCD and KITE)	Complete
July 8 2014	Multi-Modal concept	COP/IS (TCD and HON)	Complete
July 10 2014	Training concept	COP/IS (TCD and NLR)	Complete

Table X. COP Schedule of Activities: Year 2

Schedule	**Topic**	**Who Involved**	**Status**
March 3 2015	Validation of Findings of VC1 with External Pilots	COP/ES	Complete
March 25 2015	Harmonisation of Requirements Post VC1	COP/IS – All	Complete – stage 1
March 10 2015	Quantification of Bowties (evaluation of impact)	COP/IS (TCD and NLR)	Complete – stage 1

Appendix 2: Features of COP in the A-PiMod COP

Table 3. Features of COP in the A-PiMod COP

Characteristics of the 'Community of Practice' (proposed by Wenger)		
Feature	**Definition**	**A-PiMod 'Community of Practice'**
Domain	Specifies the identity of COPs with the specific competence and commitment the stakeholders engage	Involvement ofstakeholders who share technical knowledge and skillsassociated with relevant functions in the aviation-related and Flight operational system. Members will engage ina range of validation/evaluation activities on a continuous/regular basis, through the run-time of the project.
Community	Stakeholders build their relationship interacting in joint activities, sharing information and common objectives and learning from each other	Stakeholders will be involved in the project validation activities to support the progressive specification and review of requirements of the A-PiMod concept and the evaluation of associated technology prototypes.Validation sessions willruntime of the project.
Practice	Stakeholders share a repertoire of resources (experiences, stories, tools, ways of addressing recurring problems) which help forming the practice with time and sustained interaction	The panel of experts in COP will recruit internal stakeholders (i.e. representatives from each project partner) and external stakeholders. Both will contributeexpertise from theFlight operational system and other domains of intervention.

Appendix 3: Types of Sessions and Techniques

Table 4. Types of Session undertaken and different techniques used in each

COP Session	Type of Session						
	Web Mediated	In-person	Semi Structured Interview (in person or using telephone)	Web mediated participatory Workshop – involving feedback template completion and review	In Person Participatory workshop involving Co-design activities & Questionnaire	In Person Workshop	Audio Recordings
1	**X**			**X**			
2	**X**			**X**			
3	**X**	**X**	**X**				
4	**X**			**X**			
5	**X**			**X**			
6		**X**				**X**	
7	**X**			**X**			
8		**X**				**X**	
9		**X**	**X**	**X**			
10	**X**			**X**	**X**		
11	**X**			**X**			
12	**X**			**X**			
13		**X**					**X**
14		**X**			**X**		**X**
15		**X**			**X**		

Appendix 4: A-PiMod Coding Frame

Table 5. Breakdown of the A-PiMod Coding Frame

CATEGORY	**#**	**Sub-category**
ATM System (Level 0)	1.1	Operational Concept. Current and Future
	1.2	Stakeholders: Roles/responsibilities and tasks. Current and Future
	1.3	Technologies and information flow. Current and Future
	1.4	Flight Operations Process
Problem definition (Automation)	2.1	Automation current state-of-the-art
	2.2	Accident literature and automation CIT
A-PiMod concept (Automation Solution)	3.1	Overall objectives/goals
	3.2	Overall Adaptive Automation concept/philosophy
	3.3	How it works
	3.4	Roles/responsibilities: Pilot & Automation
Technology Overview	4.1	Architecture
	4.2	Components (high level) (List of main components)
	4.3	Multimodal concept MM Interaction HMMI Interaction Manager
Scenarios	5.1	High level Narrative
	5.2	Scenarios features
Underlying Concepts	6.1	Crew Model
	6.2	Workload
	6.3	Workload Measurement
	6.4	Situation Awareness
	6.5	Situation Assessment
	6.6	Risk assessment
	6.7	Task Distribution
Training	7.1	Introduction CBT
	7.2	Preparation CBT
	7.3	Instructor's Training Tool

Appendix 5: Human Factors Problems to Consider

Table 6. Human Factors Problems to Consider

Type	Example
HF problems specific to automation	Poor teamwork between crew and automation (crew have poor understanding of automation status) • Loss of situation awareness in relation to automation status – specifically, flight mode awareness, automation level awareness, impending changes of control loops and workload distribution • Poor understanding of automation modes more generally (many different mode states, confusing terminology)
	Poor teamwork between crew and automation (automation not understanding crew intentions) • Sometimes Pilot knows more than automation – experience to know can continue with flight plan and not impacted by thunderstorm • As such, issue not just whether crew understand automation, but whether automation acting in way that understand intentions of Pilot
	Overconfidence and getting lazy – danger/problem if not keeping track of status of automation
	Over-reliance on automation and loss of situation awareness
	Performance drops when not enough workload – need to be involved
	Pilot deskilling/degrading flying skills due to over-reliance on automation
	Automation not provide context information (except for Windshear and Terrain awareness)
	Lack of standardization across different aircraft types
More general HF/operational problems – that might be addressed by improved automation design	High workload – need task support
	Crew composition – gaps in experiences levels
	High workload and loss of situation awareness
	Fatigue – need task support
	Emergency situations – need task support
	Poor crew CRM
	Low workload – not involved, easily fall out of loop
Existing technology/information gaps in terms of cockpit design – that might be addressed by improved automation design	Currently no risk assessment information provided – useful to obtain this informationand associated decision options
	Currently, no crew state monitoring
	Currently, no detection of degraded crew performance – only post hoc analysis (i.e. Flight Data Monitoring)
	Currently, no monitoring of quality of teamwork between crew i.e. detecting degraded CRM
	Lack of information integration with company – Flight Operations Control
	Poor weather information – predictive information re weather and associated risk assessment

WORKPLACE RISK ASSESSMENT IN IRELAND:
IS THERE A STANDARD PRACTICE?

Hrymak V.,[1] DeVries J.M.A.,[2] Leva M.C.[3]

1; School of Food Science & Environmental Health, Dublin Institute of Technology.

2; School of Nursing and Midwifery, Trinity College Dublin.

3; Centre for Innovative Human System School of Psychology Trinity College Dublin.

Abstract

The practice of risk assessment remains problematical as there is no consensus on the hazard identification methods to be used during workplace inspections. An exploratory study on this issue, was conducted by telephone interviews with experienced risk practitioners (n=40). They were asked to describe their typical risk assessment conduct for a workplace scenario presented by the researcher (n=32) or a scenario of their own choice (n=8). The results demonstrated that while practitioners generally reported a consistent approach across their risk evaluation tasks, there was a great deal of variability among them in terms of hazard identification methods used. They all reported using an overarching paradigm of "look ask read" but there were significant variations in time taken, areas looked at, questions asked, staff consulted, documents read, and equipment used. These findings are discussed in relation to the issue of mis-identification of workplace hazards.

Introduction

Risk assessment is a legal requirement for all Irish workplaces. HSA (2015). It is a precursor to the control of workplace hazards and consists of two stages; hazard identification followed by risk evaluation. There are many risk assessment methods documented and a comprehensive listing can be found at Gould *et al*, 2005; IEC/ISO, 2009; Marhavillas *et al* 2011; Tixier *et al* 2002. The importance of the hazard identification stage cannot be understated. As Aven (2011) points out, an unidentified hazard cannot be appropriately controlled. At its simplest, hazard identification involves observation, asking questions and reading relevant documents (HSA, 2015; HSE, 2013; Neathey *et al*, 2006). However there is no consensus as to the composition or the conduct of hazards identification methods. Commentators have pointed out (Aven, 2011;

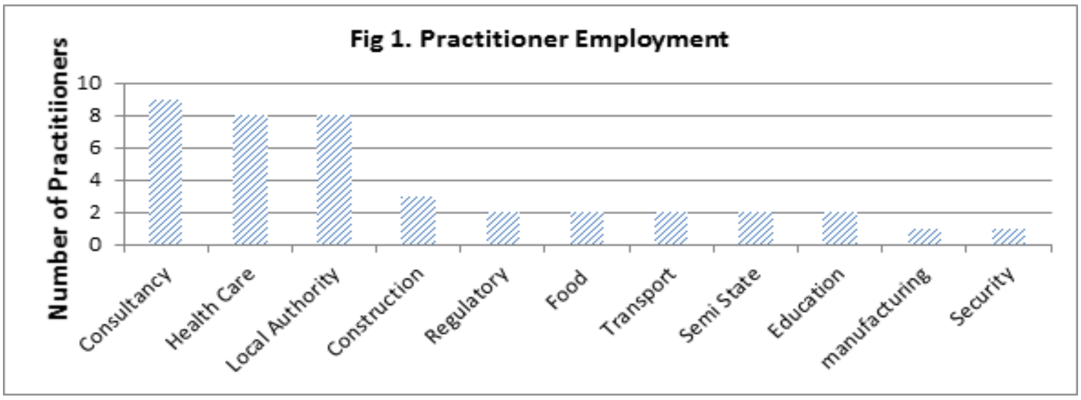
Carter and Smith, 2006; Gould *et al.,* 2005) that this lack of consensus implies that the extent to which workplace hazards are identified, can vary according to the reliability and validity of the risk assessment methods used by the individual. This concern motivated the study, which set out to explore the methods experienced Irish risk practitioners would typically use as part of their workplace inspection procedures. For clarity, the term "inspection" in this paper is the time spent on site by the practitioner, when engaged in hazard identification.

Method

The researchers recruited 40 experienced workplace risk assessors as participants, (22 Male and 18 Female). In terms of experience, the mean number of years that participants were involved with risk assessment practice was 10.61 years (SD = 6.25). In terms of professional qualifications, 39 practitioners had a degree or higher in a relevant subject, and 31 practitioners had relevant professional membership. Further descriptive statistics, are given in Table 1 and Fig 1 below.

Table 1. Descriptive Statistics of the Sample

Frequencyo f Assessments	**N Practitioner s**	**Professional Discipline**	**N Practitioner s**	**Age**	**N Practitioner s**
Daily	7	Risk	32	Twenties	5
Weekly	25	Engineer	6	Thirties	13
Monthly	7	Construction	1	Forties	15
Quarterly	1	Design	1	Fifties	7



Using telephone interviews, risk assessment methods used by practitioners were investigated. Practitioners were asked to apply their usual method to a workplace scenario of their own choice or a scenario suggested by the researcher. The researcher suggested scenario was a three to four storey building with 200-250 occupants, technically and operationally complex, similar to a financial services centre, an IT centre, a health care provider, a regional administrative or governmental head office. Most practitioners (32 out of 40) opted for the researcher suggested scenario. The remaining scenarios chosen by practitioners were 5 large construction sites, a small construction site, an engineering workshop and a food manufacturing organisation. Following their choice, practitioners were asked to apply their typical risk assessment practice and consider their response according to the following semi structured questions:

- What preparations do you make?
- What items or equipment do you take to the site?
- What questions do you ask, who do you ask and what documents do you read?
- In what order do you look at the site and how long do you spend on site?
- How are your hazards evaluated in terms of risk?

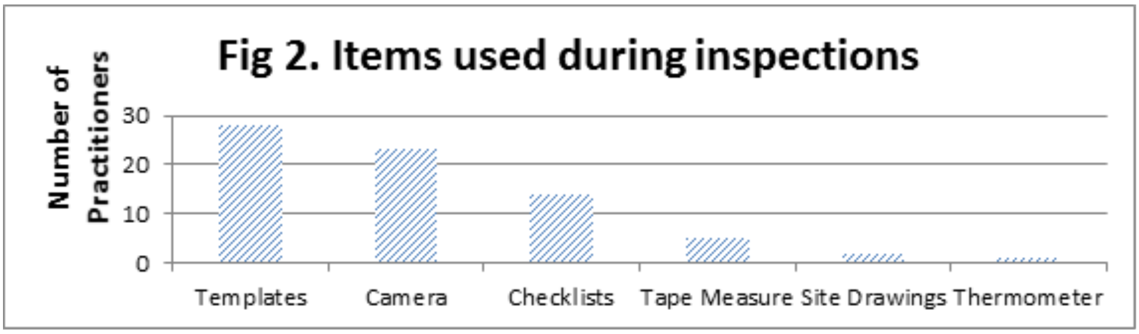
The limitations of this methodology are acknowledged. The nature of self-reported behaviour has been well critiqued (for example see Aronson, 2013). The distinction between the scenario approach used, and actual practitioner behaviour is therefore well understood. Furthermore, the small sample size is not sufficient in terms of statistical power (Cohen, 1992) and cannot represent the wider risk practitioner community. This study is best described as an exploratory study, and does not claim any statistical significance. However, this study remains a credible account of participant knowledge, and to an extent, behaviour. It is a first step in establishing what may constitute representative conduct of risk practitioners in Ireland.

Results

Pre inspection procedures were found to be standardised. All the participants demonstrated a high level of prior knowledge of the hazards to be expected from workplaces. All practitioners reported consulting any available information prior to site visits. In effect, all practitioners showed evidence that they came to the workplace with a great deal of background knowledge as to the hazards to be expected. During inspections, all 40 practitioners reported that they used three distinct hazard identification methods; observation, questioning staff and reading relevant documents. This “look ask read” approach, was clearly and unambiguously reported by all 40 practitioners. However beyond this overall approach there was very little uniformity among the participants, regarding what to look at, what order to use when inspecting the building, what questions to ask, who to ask, what documents to read, what equipment to use or how long to stay on site. These variations appeared to be personal preferences rather than task differences.

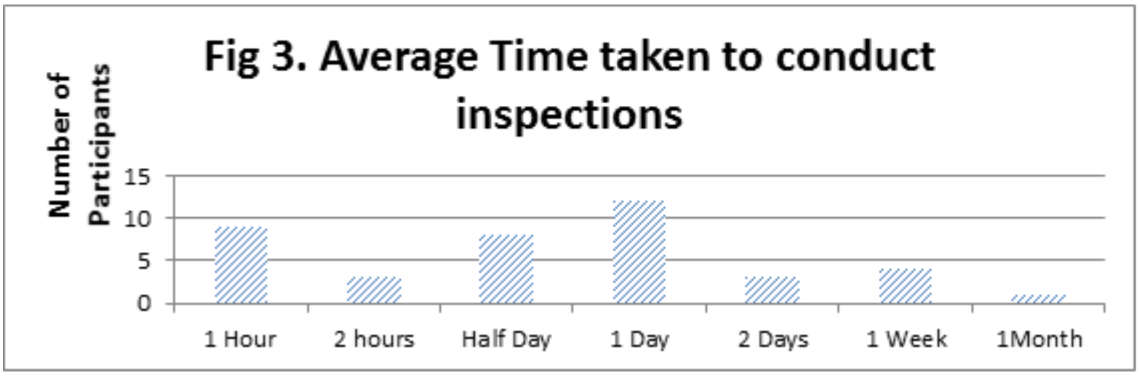
Items used during inspections

What the practitioners took with them during inspections varied. In descending order, the most frequently reported items used during inspections were; templates, cameras, checklists, tape measures, site drawings and in one case a thermometer. (see Fig. 2 below)



Time taken to conduct inspections

The variability in time taken when conducting workplace inspections was particularly large. The range reported was between one hour and one month. The mode and median time taken to conduct inspections was one working day. Twenty practitioners (50 %) reported taking less than half a day and 9 practitioners (22.5%) reported taking 1 hour to conduct inspections.



Lack of systematic or uniform hazard identification methods

Although task analysis or process mapping was used by 6 practitioners (15%), only 1 practitioner (2.5%) stated that they would observe workplaces using a set order of access. Only 2 practitioners (5%) stated they used the same questions during inspections. Only 1 practitioner (2.5%) stated they would always ask for specific individuals to interview. The remaining practitioners all reported that questions were prompted or flowed from site conditions and staff encountered rather than by specific questions. Three practitioners (7.5%) did report using "what if" questions but again there were no set questions.

The highest level of standardised inspection conduct related to document analysis. Only one practitioner (2.5%) reported specifically requiring the safety statement, the site safety file and the most recent risk assessment to be produced during the site visit. There were nine practitioners (22.5%) that reported looking for similar safety related documents but only if they were present. The remaining practitioners reported looking for safety related documents but did not specify the type of document other than to "look at what is available". In contrast to the variability in conduct during inspections, there was a high degree of standardisation when practitioners had finished and were describing the level of risk from their identified hazards. Here, practitioners used one of three risk descriptors; Likert scale low medium or high; safe or unsafe, or matrix tables with probability consequence ratings. The large majority of 40 practitioners (75%) used a risk matrix to assess the hazards identified. The matrix descriptions all involved the multiplication of probability and consequence ratings to produce risk levels, as per consensus guidance (e.g. BS OHSAS 18001, 2007).

Discussion

The small sample size of 40 practitioners precludes generalisation to a wider population and restricts the data from any inferential statistical treatment. However the homogeneity of the findings from this exploratory study, points to the possible, if not probable, wider professional existence of an overarching "look ask read" approach during inspections which is accompanied by large variation in hazard identification conduct. This variation was particularly noteworthy in light of the fact that 80% of practitioners responded to the same scenario. However, given the limitations of this study, any interpretation must be cautious. With this warning in mind, the discussion here is restricted to a consideration of how this variation has arisen, and how it can be resolved in the interest of workplace safety.

The main finding in this study remains the wide variation within the "look ask read" approach to conducting workplace inspections. The practitioners in the sample were experienced, highly educated and competent in terms of professional body membership. So why was practitioner conduct during inspections so varied. On reflection, there are a number of possible factors to explain this. However it is not really surprising that chief amongst these explanations is the lack of consensus and guidance as to what constitutes adequate inspection procedures for occupational risk assessments.

Although the call to improve risk assessment has been repeatedly made, (Aven, 2011; Aven, 2013; Gould *et al,* 2006; Pinto *et al,* 2013) there remains little consensus regarding standardising the practice within the risk assessment community. There is a legal requirement to identify all "reasonably foreseeable" workplace risks but as Aven (2011) points out, we all know what risk is, but there is no consensus as to how to measure it. A reflective practitioner in Ireland asking the question, what is the best way to conduct a workplace risk assessment?, would find little in the way of definitive guidance. It should be noted that there is a lack of consensus for good reason. As Le Coze, (2005) points out, the conduct of an appropriate risk assessment is a delicate balance. If the method is too simplistic, it will not have validity. If the method is too complex it becomes time consuming, requires a technical professional and will quickly become uneconomic. There are examples of standardisation available. As pointed out by Leva et al, (2015) task analysis based risk assessment methods for safety critical operations include HAZID; (EN ISO, 17776:2002) job safety analysis (US Dept of Labor, 2002) and Facility Hazard Analysis (Taboas et al, 2004). The UK based Royal Institute of Chartered

Surveyors (RICS, 2010) also have very specific procedures when inspecting buildings. However there is still a lack of wide consensus on actual inspection and risk assessment procedures to adequately identify the wide spectrum of workplace hazards.

The professionalism of the practitioners who participated in this study can only be complemented. However it is also apparent, that there is no external pressure to reflect on the need for improvements in hazard identification methods. In terms of the wider workplace safety community, the concern is that hazards detected are predicted more by the conduct, diligence or thoroughness, of the inspector, and not by the presence of actual hazards within that workplace. Taking the example of time spent on inspections, (and acknowledging the premises size differences in the scenarios) the variability is so wide that there can be little consistency amongst practitioners. Simply put, an inspector spending one hour on a large site, is highly unlikely to identify the range of workplace hazards that must exist, when compared to an inspector who chooses to spend a full day on the same site. Standardisation in hazard identification methods during inspections would therefore, greatly benefit the risk assessment process. The observation of hazards during inspections remains the cheapest, quickest, most effective as well as the most widespread inspection method available. Therefore standardisation of observational hazard identification practice, is particularly important.

This renewed call for the standardisation of hazard identification methods for workplace inspections is relatively easy to make. In comparison, the research required to provide an evidence base that will support effective and appropriate inspection conduct, is far more challenging. The researchers have begun work on this evidence base, and are currently involved in inspection behaviour research, including observational hazard identification capability. Early results, suggests that humans are highly error prone when tasked with inspecting workplaces for hazards. However, this field based applied research has also found that specific visual training, can significantly increase the observational hazard identification capability of inspectors. Currently, applied research into effective inspection behaviour remains largely unexplored and does not reflect the humanitarian and economic importance of workplace safety. This study, reinforces the need for better guidance, a more standardised approach to risk assessment and as well as the rigorous evaluation of methodological effectiveness when identifying workplace hazards.

Acknowledgement

The researchers would like to recognise the role of Irish risk practitioners in this study. In particular, their professionalism and high standards were clearly evident. This study should not be seen as criticism. Instead, their desire to self-improve and the reflective nature of these professionals, is clearly evidenced by their co-operation in this study.

References

Aronson, E., 2013. *The Social Animal.* Worth Publishers.

Aven, T., 2001. *Selective critique of risk assessments with recommendations for improving methodology and practice Reliability Engineering and System Safety* **96**, 509-514.

Aven, T., 2011. *Quantitative Risk Assessment, the Scientific Platform*. Cambridge University Press.

BS OHSAS 18001., 2007. *Occupational health and safety management systems requirements.* British Standards Institute.

Carter, G., Smith S.D., 2006. *Safety Hazard Identification on Construction Projects. Journal of Construction Engineering Management.* **132**:197-205.

Cohen J., 1992. *A Power Primer.* Psychological Bulletin, 1992, **Vol. 112**. No. 1,155-1,159

EN ISO 17776:2002 Petroleum and natural gas industries —Offshore production installations —Guidelines on tools and techniques for hazard identification and risk assessment.

Gilovich, T., Griffin, D.W.H., Kahneman, D., 2013. *Heuristics and Biases, the Psychology of Intuitive Judgement.* Cambridge University Press.

Gould, J., Glossop, M., Ioannides, A. 2005. *Review of Hazard Identification Techniques. HSL/2005/58.* Health and Safety Laboratory, HMSO. UK.

HSA, 2015. *SAFETY REPRESENTATIVES RESOURCE BOOK*. Health and Safety Authority. Dublin Ireland.

HSE 2014. *Risk Assessment A brief guide to controlling risks in the workplace.* INDG163.(rev4) Health and Safety Executive, UK.

IEC/ISO 31010, 2009. *Risk Management - Risk Assessment Techniques*. International Standards Organisation. Geneva, Switzerland.

Le Coze J.C., 2005. *Are Organisations Too Complex For Technical Risk Assessment and Current Safety Auditing?.* Safety Science **43** 613-638.

*Leva M.*C. Naghdali F., Balfe N. Gerbec M., De michela M.2015 *Remote Risk Assessment: Case Study using SCOPE Software*. ***CHEMICAL ENGINEERING TRANSACTIONS VOL. 43.***

Marhavillas, P.K., Koulouriotis, D., Gemeni, V., 2011. *Risk analysis and assessment methodologies in worksites. J. Loss Prevention in the Process Industries* **24**, 477-523.

Neathey, F.; Sinclair, A.; Rick, J.; Ballard, J.; Hunt, W.; Denvir, A. 2006. *An evaluation of the five steps to risk assessment.* Research Report 476. Institute for Employment Studies.

Pinto, A., Ribeiro, R.A., Nunes, I.L. 2013. *Ensuring the Quality of Occupational SafetyRisk Assessment. Risk Analysis*, **Vol. 33**, No. 3. 409-419.

RICS, 2010. *Practice Standards. UK Building surveys and technical due diligence of commercial property.* 4th Edition Guidance Note GN23/2010. Royal Institute of Chartered Surveyors. UK.

Tixier, J., Dussere, G., Salvi, O., Gaston, D., 2002. *A Review of 62 risk analysis methodologies of industrial plants. Journal of Loss Prevention in the Process Industries.* **15**, 291-303.

Taboas A., Moghissi A. A., LaGuardia T S. 2004 The Decommissioning Handbook, American Society of Mechanical Engineers, Published by ASME Press US Dept of Labor 2002 Job Hazard Analysis. OSHA 3071.

THE IDENTIFICATION OF ASSESSMENT CRITERIA FOR A SAFETY REPORTING SELF-ASSESSMENT TOOL

E.Douglas, C.Leva, S.Cromie

Center for Innovative Human Systems, Department of Psychology, Trinity College Dublin, Éire

Abstract

This paper describes the assessment criteria that have been identified for use within a self-assessment tool for risk reporting within an industrial context. An in-depth review of the available literature was carried out exposing the criteria discussed within this paper. The criteria target all areas of reporting from the design of the reporting interface to the cultural considerations and initiatives. These criteria will be refined and weighted using industrial case study experience and then developed into the full self-assessment tool.

Introduction

Why Reporting is Important within an Organisation?

Learning from near misses, previous mistakes and operational experience has historically been a cornerstone within safety management for as far back as the Heinrich, (1941) hierarchy of accident severity where it was shown that for every major incident there were many cases of smaller incidents and behind those smaller incidents there were many near misses that can occur within the organisation. This concept has remained a key consideration within safety management today with more recent models such as the Carter (2006) statistical model for accident prevention that suggests that an unidentified incident has both an unknown severity and an unknown likelihood and therefore an unknown level of risk to the organisation and the process and therefore will pose an unknown risk to the organisation which may in fact be critical to the operation.

The commonly cited "Swiss Cheese" model by Reason (1998) suggests that an accident is the result of a combination of the weaknesses in defence barriers (i.e the holes in the cheese), therefore risk reporting can be seen as a methodology to identify the holes in these safety barriers. A famous example whereby risks were not adequately appreciated can be found within the Chernobyl disaster. In this accident a combination of a lack of operator experience, poor risk assessment, management pressure and poor plant design caused the disaster (Pidgeon & O'Leary, 2000). During the investigation into the accident, it was found that an almost identical situation had been mitigated at a similar power plant, the operators of the plant had reported the near miss to the atomic regulatory bodies but both regulators and plant operators did not act on their report undermining the whole safety management process (IAEA, 1991).

The lessons from incidents have driven organisations to develop and implement safety management systems (SMS). SMS systems take data that has been inputted from a variety of sources that can include reports submitted from staff within the organisation. These reports can pertain to concerns for safety, suggestions in addition to near misses and incidents. In order to improve the amount of "good" data leading into the system, SMS systems are frequently citied within accident reports as recommendations that should be implemented within the organisation (Leveson, 2011). However previous studies have found that reporting systems on site are underutilised and under used (Bhattacharya, 2012; Cromie et al., 2012; Leveson, 2011) organisations need to monitor the performance of reporting in order to ensure that a high level of reporting is maintained on site within the organisation. There is a considerable amount of literature that exists that looks into the influencing factors of reporting ranging from the design of the reporting system and procedure to the cultural elements that have to be considered within the process. This guidance does not currently provide a unified best practice and methodology which is one of the key inputs of the system.

Why should reporting be assessed with an organisation?

As discussed in the previous section, reporting systems are an important element of safety management/ maintaining compliance is a key concern for all safety critical industries (Leveson, 2011) and this therefore causes simple compliance to be cited as the reason for the development of Safety management systems and in turn reporting systems. The use of regulations being the main motivational factor for reporting can introduce ramifications for a reporting system. There is a real risk that an organisation will set up a reporting system in order to "tick the box" and then just neglect to act on the data they are inputting which can result in a reporting system that isn't used within the organisation as found within previous industrial studies (Cromie et al., 2012; Kongsvik, Fenstad, & Wendelborg, 2012) where the level of reporting declined soon after the reporting system was implemented and the opportunity to use reporting to reduce the level of risk within an organisation was missed.

In Reason (2004) a safe organisation is described as an organisation whereby there is a large amount of knowledge being inputted into Safety Management. A complete lack of reports demonstrates an organisation with a low level of risk awareness. In Reason (2004) a safe organisation is described as an organisation whereby there is a large amount of knowledge being inputted into the SMS. A complete lack of reports demonstrates an organisation with a low level of risk awareness. In the statistical model for accident causation developed by Carter and Smith (2006) a hazard that is not identified can have both a fully unknown level of severity and likelihood and therefore an unknown level of risk. Therefore it is crucial for organisations to ensure the reporting system is utilised as much as possible in order to maintain a high level of safety knowledge. Despite this there is a tendency for an organization to implement a reporting system and then fail to ensure that the system is actually identifying hazards etc. (Cromie et al., 2012; Leveson, 2011), Reason (2004) argues that a low number of hazard and near miss reports are not an indicator of a safe organization but an indicator or an organization that has poor risk management. Therefore there is an opportunity to develop a methodology that will allow organizations to assess their reporting system. It is proposed to have two areas of assessment, one focusing on the individual role within reporting and a second looking at the managerial oversight for reporting:

- Individual focus of reporting – A survey methodology
- Managerial focus of reporting – A Reporting Self-Assessment Tool (R-SAT)

The R-SAT will be a software tool that will allow organizational management to assess the key influencing factors of reporting based on the assessment criteria that will be discussed within this paper.

Identification of Criteria

A high level overview of a reporting process was established. This was used to provide guidance to assist in investigating the existing literature and case studies to distil the main requirements and objectives of reporting systems and use them as the basis of the high level evaluation criteria. The criteria identified through this process are shown in Table 1 and a discussion on how and where they were identified in the literature is reported in the following chapter. Once these categories were established an in-depth literature review was carried out and the individual assessment criteria were developed:

Table 1. Assessment Criteria

Evaluation Criteria	Sub Criteria
Meeting the Requirements of the Regulatory Environment	Regulatory Requirements
	"Good Practice"
Usability of reporting forms and feasibility of reporting Procedure	Reporting Form Design
	Reporting Procedure
Provision and value of Feedback	Feedback loop To Reporters
	Value of Feedback for the organization
	Value of good catches to the organisation
Education and Promotion of the Reporting System	Training
	Reporting Awareness
Motivation	Safety Culture
	Stimulation of Reporting
External Influences from the Industrial sector	External factors

Identification of Assessment Criteria

Meeting the Requirements of the Regulatory Environment

The Regulatory Environment will vary depending on the industry that is being assessed. Typically safety critical industries such as Aerospace and the Process industry have fairly strict regulations governing the use of reporting systems within Safety Management. To develop and validate the R-SAT it is proposed to develop and validate two regulatory assessment criteria, one targeting the (Civil Aviation, 2013; European Aviation Safety Agency, 2013) and the other targeting the European Commission (1994) regulations of pharmaceutical good manufacturing practise regulations.

Table 2. Aerospace Reporting Requirements

Reference	Criteria
4.5.3	The Organisation has a SMS Manual which communicates the SMS policy to the whole organisation and provides the following: Scope of the SMS Objectives of the SMS Hazard Identification and Risk Management schemes Incident Investigation and Reporting plans Just Culture policy and Culpability Definitions are propagated to staff The SMS is promoted on site
5.1	The organisation has a process of on-going hazard identification
5.1	The organisation has a process of hazard identification and reporting using a mixture of reactive and proactive approaches including safety surveys, near miss reporting, hazard reporting systems etc.
6.1.1	Safety audits are implemented to ensure that the SMS is sound in terms of: Adequate staff levels Levels of competency Achievement of Safety Policy and Objectives Effectiveness of interventions and risk mitigations
6.1.2	Safety and Cultural surveys are used to examine the effectiveness of a specific operation and can use: Checklists Questionnaires Interviews

Reference	Criteria
6.3	The Organisation seeks to improve SMS performance, continuous improvement could be achieved through: Proactive and reactive evaluation of day-to-day operations though safety audits or surveys Evaluation of an individual's performance to verify the fulfilment of their safety responsibility Change Management
7.2	Safety Communication, the SMS is propagated by : Policies and Procedures Newsletters, safety bulletins etc. Websites and e-mail Informal workplace meetings etc.
7.2.1	Staff should be familiar with the SMS and the Organisational Safety Culture: Disseminate safety critical information Feedback explains why actions are taken

Table 3. GMP Regulation

Reference	**Criteria**
1.8 (vii)	Any significant deviations are fully recorded, investigated with the objective of determining the root cause and appropriate correction and preventative action implemented
1.10 (iv)	A review of all significant deviation or non-conformances, their related investigations and the effectiveness of resultant corrective and preventative actions taken
1.12	Quality risk management is a systematic process for the assessment, control, communication and review of risks to the quality of the medicinal produce. It can be applied both proactively and reactively.

Usability of reporting forms and feasibility of reporting Procedure

This concerns the design of the reporting form and the procedures surrounding reporting. There has been previous research into design considerations for the reporting form. Shown below in

Table 4 are the assessment criteria for the form design, considering the ethos of reporting, design considerations based on the referenced literature

Table 4. Reporting System Design

Reference	Criteria
Kongsvik, Fenstad, & Wendelborg, (2012) Leva et al. (2010) Evans et al., (2006).	The reporting system is implemented with a view to reduce additional paperwork and ideally should aim to reduce the paperwork reporting staff
Boeing Airplane (2000),Leveson (2011)	The reporting form allows reporters to share information that focuses outside of the act of the error including influencing factors such as environmental, workload etc.
Williamsen (2013)	The form is constantly reviewed for its suitability for purpose
Lappalainen et al., (2011)	Efforts are made to ensure that the reporting form is clear concise and to the point
(Johnson, 2002; Leva, Cahill, et al., 2010)	Any electronic reporting forms that are developed are intended to be easily accessible
Leva et al. (2010) (Lappalainenet al., 2011)(DePasquale & Geller, 2013; Short & Keasey, 1997)	The reporting system is designed in such a way to require a minimum amount of time to submit a report. The Organisation monitors the time impact report submission has on reporting staff.
Cohen (2000) Dekker (2012)	Reports are encouraged to report not through mandatory approaches but out of interest for safety. Management actively reinforces this through the reporting procedure.
Dekker (2012)	The organisation is moving to a voluntarily reporting system if they are already not at the point.

Provision and value of Feedback

Feedback is commonly cited as one of the key motivational factors towards reporting (S. M. Evans et al., 2006; Sue M. Evans et al., 2007; McAfee & Winn, 1989; Reason, 1998; Sanne, 2008) therefore these criteria will be looking at the provision of feedback within the organization.

Reference	Criteria
(S. M. Evans et al., 2006; Sue M. Evans et al., 2007; McAfee & Winn, 1989; Reason, 1998; Sanne, 2008).	The Organisation laid out procedures for the delivery and objectives of feedback within the organisation
Williamsen, (2013)	The quality of feedback within the organisation ismonitored by management
(Dekker & Stoop, 2012)	If a mitigation isn't available or feasible, then steps are taken to explain to the reporter why
(Douglas, Cromie, Leva, & Balfe, 2014)	There is a feedback loop for reporting within the organisation

Education and Promotion of the Reporting System

These criteria concerns the efforts implemented towards educating reporters and promotion initiatives of the reporting system within the organization. These criteria will focus on the operational.

Reference	Criteria
(Erdoğan, 2011; Leveson, 2011)	The Organisationhas clear concise guidelines for the use of safety reporting within industry
(Baram & Schoebel, 2007; Krugh & Sommers, 2010; Storgård, Erdogan, Lappalainen, & Tapaninen, 2012)	The Organisation makes efforts to make staff aware of the importance ofreporting on site, and avoids target based approaches
(Cohen, 2000)	The organisation has a voluntarily approach to reporting or plans to move to a voluntary approach
(McAfee & Winn, 1989)	Positive mitigations are used as an example to encourage more reports
Douglas et al. (2015)	There are efforts to remind staff about the existence of and benefits to the reporting system

Motivation

These criteria look into the motivational techniques management can implement to stimulate reporting on site focusing on the cultural considerations.

Reference	Criteria
(Reason, 1998; Waring, 2005; Williamsen, 2013)	The organisation has a clear culpability agreement that makes staff aware of what types of behaviour and reports are acceptable or not
Reason, 1998; Waring, 2005;Clarke (1998b)	Reports are seen as simply a way to improve safety and not a way to catch people out for unsafe behaviour
Reason, 1998;	The Organisation implements an approach to advertise positive mitigations as a result of the SMS system
Waring, 2005;(Clarke, 1998a; Sue M. Evans et al., 2007; Walton, 2006)	Efforts are made to break the seniority gradientwithin the organisation so that junior staff have the confidence to report safety concerns and near misses involving more senior staff.
(Smith, 1999)	The organisation is avoiding Behavioural Based Safety approaches and is instead trying to help instil the attitude that reports improve the safety of someone's working environment and can provide a net benefit to their own safety

External Influences on the Organization

These final criteria look at the role of external factors on the organization. These factors can range from recent incidents within the industrial sector, litigation problems such as Freedom of Information requests. These factors will not be under influence of the management but they can have an effect on the reporting system performance.

Reference	Criteria
Waring, 2005;(Sue M. Evans et al., 2007)	Efforts are made to protect reports from litigation such as freedom of information effects
(Leveson, 2011)	Accidents within the industry are discussed within the organisation and

	report recommendations can be used as a driver for reporting
(Reason, 1998)	The reporting system is developed with cooperation by unions and other interest lobbies

Further Work and Conclusions

The assessment criteria discussed in this paper will be further refined using case study experience in the Biopharma and Aerospace industries where surveys and semi structured interviews have been carried out to help expose additional areas of assessment. In addition the case studies will allow the weighting of the assessment criteria to be determined to help assist in the development of a quantitative assessment of the organizational reporting culture which combined with techniques such as dashboards etc. will allow a safety manager to drill down and determine the areas that will need attention and improvement.

References

Baram, M., & Schoebel, M. (2007). Safety culture and behavioral change at the workplace. *Safety Science, 45*(6), 631-636. doi: 10.1016/j.ssci.2007.04.001

Bhattacharya, S. (2012). Sociological factors influencing the practice of incident reporting: the case of the shipping industry. *Employee Relations, 34*(1), 4-21. doi: 10.1108/01425451211183237

Boeing Airplane, C. (2000). Maintenance Error Decision Aid (MEDA) - Users Guide. Seattle WA.

Carter, G., & Smith, S. D. (2006). Safety Hazard Identification on Construction Projects. *Journal of Construction Engineering and Management, 132*(2), 197-205. doi: 10.1061/(ASCE)0733-9364(2006)132:2(197)

Civil Aviation, A. (2013). CAP 716 Aviation Maintenance Human Factors (EASA / JAR145 Approved Organisations).

Clarke, S. (1998a). Organizational factors affecting the incident reporting of train drivers. *Work & Stress, 12*(1), 6-16. doi: 10.1080/02678379808256845

Clarke, S. (1998b). Safety culture on the UK railway network. *Work & Stress, 12*(3), 285-292. doi: 10.1080/02678379808256867

Cohen, M. R. (2000). Why error reporting systems should be voluntary. *BMJ, 320*(7237), 728-729. doi: 10.1136/bmj.320.7237.728

Cromie, S., Liston, P., Ross, D., Corrigan, S., Vani, L., Lynch, D., & Demosthenous, S. (2012). Evaluation report on Phase 1 of the (The Company's) Human Factors and Safety Management Programme 2011-2012: Trinity College Dublin. Dublin: Trinity College.

Dekker, S., & Stoop, J. (2012). Are safety investgations pro-active? *Safety Science, 50*, 1422-1430.

DePasquale, J. P., & Geller, E. S. (2013). Reprint of “Critical Success Factors for Behavior-Based Safety: A Study of Twenty Industry-wide Applications”. *Journal of Safety Research, null*(null). doi: 10.1016/j.jsr.2013.07.007

Douglas, E., Cromie, S., Leva, M. C., & Balfe, N. (2014). Modelling the Reporting Culture within a Modern Organisation. *Chemical Engineering Transactions, 36*, 589-594. doi: 10.3303/CET1436099

Erdoğan, I. (2011). *Best Practices in near-miss reporting.*

European Aviation Safety Agency. (2013). EASA Part 145 2013-01 (Vol. 0251).

European Commission. (1994). Directive 94/9/EU of the European Parliament and Council. *Official Journal of the European Union*(100), 1-7.

Evans, S. M., Berry, J. G., Smith, B. J., Esterman, A., Selim, P., O'Shaughnessy, J., & DeWit, M. (2006). Attitudes and barriers to incident reporting: a collaborative hospital study. *Quality & safety in health care, 15*(1), 39-43. doi: 10.1136/qshc.2004.012559

Evans, S. M., Smith, B. J., Esterman, A., Runciman, W. B., Maddern, G., Stead, K., . . . Jones, S. (2007). Evaluation of an intervention aimed at improving voluntary incident reporting in hospitals. *Quality & safety in health care, 16*(3), 169-175. doi: 10.1136/qshc.2006.019349

Krugh, K. a., & Sommers, J. R. (2010). Safety Observation System and Near Miss Reporting Programs: Methods to maintain employee engagement in behavioral based safety practices. *2010 IEEE-IAS/PCA 52nd Cement Industry Technical Conference*, 1-10. doi: 10.1109/CITCON.2010.5470071

Leveson, N. G. (2011). *Engineering a Safer World*. Cambridge ,MA: Massachusetts Institute of Technology.

McAfee, R. B., & Winn, A. R. (1989). The use of incentives/feedback to enhance work place safety: A critique of the literature. *Journal of Safety Research, 20*(1), 7-19. doi: 10.1016/0022-4375(89)90003-0

Reason, J. (1998). Achieving a safe culture: Theory and practice. *Work & Stress, 12*(3), 293-306. doi: 10.1080/02678379808256868

Reason, J. (2004). Beyond the organisational accident: the need for "error wisdom" on the frontline. *Quality and Safety in Health Care, 13*(suppl_2), ii28-ii33. doi: 10.1136/qshc.2003.009548

Sanne, J. M. (2008). Incident reporting or storytelling? Competing schemes in a safety-critical and hazardous work setting. *Safety Science, 46*(8), 1205-1222.

Short, H., & Keasey, K. (1997). Institutional Voting in the UK: Is Mandatory Voting the Answer? *Corporate Governance: An International Review, 5*(1), 37-44. doi: 10.1111/1467-8683.00038

Smith, T. A. (1999). What's wrong with behavior based safety ? *Professional Safety*(September), 1-8.

Storgård, J., Erdogan, I., Lappalainen, J., & Tapaninen, U. (2012). Developing Incident and Near Miss Reporting in the Maritime Industry–A Case Study on the Baltic Sea. *Procedia - Social and Behavioral Sciences, 48*, 1010-1021. doi: 10.1016/j.sbspro.2012.06.1078

Walton, M. M. (2006). Hierarchies: the Berlin Wall of patient safety. *Quality & safety in health care, 15*(4), 229-230. doi: 10.1136/qshc.2006.019240

Waring, J. J. (2005). Beyond blame: cultural barriers to medical incident reporting. *Social science & medicine (1982), 60*(9), 1927-1935. doi: 10.1016/j.socscimed.2004.08.055

Williamsen, M. (2013). Near-Miss Reporting - A Missing Link in Safety Culture. *Professional Safety*(May).

SEDENTARY BEHAVIOUR IN THE CLASSROOM: A COMPARISON BETWEEN GERMAN AND IRISH SECONDARY SCHOOLS

Burns J,[*] Breithecker D,[§] Mahli M.[§] Gormley J.[*] Dockrell S.[*]

[*] *Department of Physiotherapy, University of Dublin, Trinity College, Dublin 2*

[§] *Bundesarbeitsgemeinschaft für Haltungs- und Bewegungförderung e.V., Kirchkohl 14, 65207 Wiesbaden Naurod, Germany*

Abstract

A classroom-based observational study was undertaken to investigate sedentary behaviour in Irish and German secondary school children. A total of 29 children were observed for one day each and their behaviour was recorded at 1-minute intervals using an observation checklist. The children spent the majority of the time sitting (89.8%), although 76.5% of that time was spent in 'active sitting'. German children had significantly more 'active sitting' than Irish children ($p<0.001$). There was no gender difference in the classroom behaviour ($p=0.81$). The classroom behaviour varied significantly between practical and non-practical subjects ($p<0.05$). Sedentary behaviour is a modifiable risk factor for many conditions and as such, from a public health point of view, it is important to quantify and analyse the time spent in sedentary behaviour in the classroom.

Introduction

The increase in sedentary behaviour in children has been of concern for some time (World Health Organisation, 2014). It is estimated that children (6-11 years) spend on average six hours a day engaging in sedentary behaviour while adolescents (16-19 years) spend approximately eight hours a day in sedentary behaviour (Matthews et al, 2008). The definition of sedentary behaviour used in the literature has not been consistent. Typically in sports and exercise literature, sedentary behaviour has been used to describe insufficient or a lack of moderate to vigorous physical activity (Mullen et al, 2011; Sims et al, 2012; Smith et al, 2010). However in 2012 the Sedentary Behaviour Research Network (SBRN) proposed that sedentary behaviour be defined as, '*any waking behaviour characterised by an energy expenditure ≤1.5 METs while in a sitting or reclining posture'* (SBRN, 2012).

Although sedentary behaviour and physical activity are closely linked, sedentary behaviour is proposed to be an entity in itself and not simply insufficient physical activity. It is proposed that a lack of or insufficient physical activity has a different physiological effect on the body than that of sedentary behaviour (Hamilton et al, 2007; Tremblay et al, 2010). In addition, the responses and adaptations of the body to sedentary behaviour may not necessarily be the opposite to those of physical activity (Tremblay et al, 2010). Furthermore, a person can be both active and sedentary. People who meet or even exceed the recommended daily exercise guidelines, but sit eight or more hours are both active and sedentary. The prolonged sedentary behaviour during the day can have negative health consequences, regardless of the benefits gained by participating in physical activity (Owen et al, 2000).

Previous research on sedentary behaviour in children has focussed largely on leisure time i.e. outside of school hours (Tremblay et al, 2011A; Arundell et al 2013) and the extent and pattern of sedentary behaviour during school is not well documented. School-based studies have addressed activity levels during break-time (Ridgers et al, 2012; Nettlefold et al, 2011) and also during PE class (Rowe et al, 2004; Nettlefold et al, 2011). However children spend a considerable amount of time in a classroom, therefore the levels of sedentary behaviour within the classroom need to be examined. Treuth et al (2007) included school and leisure time in their study of physical activity in adolescent girls, but they did not differentiate between weekday school and leisure time. Other studies of physical activity in the classroom, measured by accelerometry, reported on sedentary behaviour (Donnelly et al, 2009; Nettlefold et al, 2011). A recent school and home-based observational study on a small sample of primary school children (n=9) documented the time engaged in different tasks with a particular emphasis on IT use, but did not include sedentary behaviour (Ciccarelli et al, 2011).

Sedentary behaviour in children has been measured using various subjective and objective methods. Direct observation allows for contextual information to be gathered, and has a high inter-rater reliability ($r>0.9$) (DuRant et al, 1994) and was used in the current study. The overall aim was to establish the level of sedentary behaviour in the classroom in a typical school day. The objectives were to calculate the time spent in different behaviours, and to investigate the effects of gender, nationality, and class subject on sedentary behaviour using an observation checklist. The study also established the tasks and sitting positions of the children in the classroom.

Method

Research Design

A school-based cross-sectional study design with an observational measurement tool was used.

Sample and Participants

A convenience sample of post primary schools in Ireland and in Germany was included in the study. Children in the participating schools were eligible for inclusion. They were excluded if they were in Transition Year.

Measures

Direct observation has been proposed as an appropriate and reliable method for the assessment of behaviour in children (Anderson et al, 1985; Du Rant et al, 1994; Rowe et al 2004), yet no previous observation checklist to observe children's physical behaviour during class time at school was found in the published literature. An observation checklist was developed for this study and contained items on demographic characteristics, behaviour (physical activity, ambulating, standing, sitting), task (write, read, listen, talk organise, make/draw, computer, think, media, other), sitting position (front seat, backrest Y/N; back seat, backrest Y/partial/N, other). Furthermore the observation checklist was used to identify, within the sitting behaviour category, the extent of movement that occurred while sitting, termed 'passive sitting' or 'active sitting' (Table 1). This was used to investigate the nature of the children's sedentary behaviour and the potential energy expenditure of the children even while they were sitting in class. The observation checklist was found to have high inter-rater reliability (r=0.9), intra-rater reliability (r=0.9), and concurrent validity when compared to an accelerometer (r=0.83).

Table 1. Descriptors of sitting behaviour

Body Part	**Passive sitting**	**Major Mov Active sitting**
Head and Neck	Movement <20°	Movements >20°
Upper Limb	hand and Hand and wrist mvt only	Including elbow and shouldermvt
Trunk	Movement <20° Any axis, minor change in WB	Movements >20° Any Axis, obvious changes in WB
Lower Limb	Ankle and footmvtonly	Including knee and hipmvt

Procedure

Schools were contacted directly by the researchers and informed of the study. If the school principal agreed to participate, information packs containing information sheets, consent and assent forms were given to the schools for distribution to randomly selected students from the school register. Once consent and assent were obtained, dates were arranged with the respective schools for the researchers to carry out the observations. On data collection days, the researcher sat discretely in the classroom ensuring a good view of the child, without disturbing the normal rhythm of the class. The observation checklist was used to record the child's movements at 1-minute intervals. Mid-morning and lunch breaks were excluded from the observations as it was deemed to be too invasive, and also because the outcome of interest was the sedentary nature of classroom behaviour.

Data processing and analysis

The observation checklist data were coded and entered into Excel for analysis. Descriptive statistics of the behaviours in class were calculated in absolute terms (minutes) and as a proportion of the total number of observations made. Descriptive statistics were used to illustrate the proportion of time spent in the behaviour categories (physical activity, ambulating, and standing, sitting) and the proportion of time spent in the sitting behaviours (active or passive). The behaviours 'physical activity', 'ambulating' and 'standing' were collapsed into a single category and labelled 'non-sitting'. The class subjects were grouped into four categories; languages, scientific, practical and academic. Independent t-tests were used to analyse the association between the mean time spent in the classroom behaviours and gender or nationality. A Pearson's Chi-square test was used to compare the proportion of time spent sitting and class subject. Statistical significance at $p<0.05$ was assumed.

Results

A total of 29 children were included in this study: 18 German children (eight boys and ten girls) and 11 Irish children (eight boys and three girls). The children were aged between 12 and 17 years old.

Table 2. Details of observations

	No. of children (n)	**Total observed time (mins)**	**Non-sitting time* (mins)**	**Sitting time (mins)**
Germany	18	4,124	459 (11.1%)	3.665 (88.9%)
Ireland	11	2,098	178 (8.5%)	1,920 (91.5%)
Total	**29**	**6,222**	**637 (10.2%)**	**5,585 (89.8%)**

*Includes standing, ambulating, and physical activity

Nationality

The children spent the majority of their time in sitting (89.8%), with little time engaged in the non-sitting classroom behaviours. There was no significant difference between the mean percentage sitting time for German and Irish children (88.9% v 91.3%; $p=0.359$). For the total population, active sitting comprised 76.5% of the observations. German children spent a significantly greater mean proportion of time ($p<0.001$) sitting actively (84.7%) compared to the Irish children (61.3%).

Gender

Of the total number of observations, 44.8% (2,789) were carried out on girls and 55.2% (3,433) were carried out on boys. There was no significant difference between the genders (89.5% v 90.1%) in the mean proportion time spent sitting in class ($p=0.81$). Girls spent a greater mean proportion of their time (80.6%) in active sitting than boys (71.9%), but the difference between them was not significant ($p=0.19$) as shown in Figure 1.

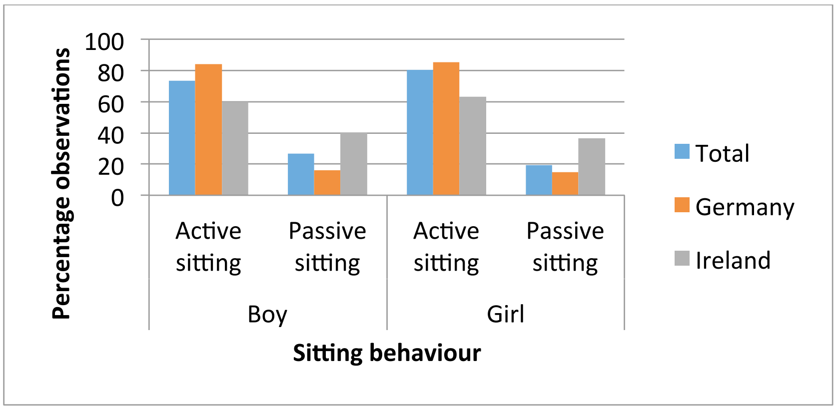


Figure 1. Type of sitting behaviour by gender and nationality

Class subject categories

There was a significant difference in the sitting time observed between the class subject categories ($p<0.05$). Similar proportions of observations of sitting were recorded in language (92%), scientific (90%) and business and social (91.2%). The proportion of sitting time in practical classes was lower (81.7%). There was also significant variation of the sitting behaviour during the different class subject categories ($p<0.05$). Active sitting behaviour comprised 89.4% of sitting time in practical classes but was lower for language (77.4%), scientific (76.0%), business and social classes (58.6%).

Discussion

It is important from a health point of view to quantify the proportion of time that children spend sitting during the school day. Sitting has been shown to be associated with back pain in children and adolescents (Balague et al, 1999; Watson et al, 2002) and prolonged sitting has recently been identified as an occupational health risk for other conditions such as diabetes mellitus, some cancers and mortality (van Uffelen et al, 2010; Tremblay et al, 2011A). There is also evidence that decreasing sedentary behaviour in children is associated with a lower health risk (Tremblay et al, 2011B). The school years are an optimal time to learn good working practices regarding sitting, as this is a behaviour that is typically carried through into adulthood and will impact on working life and leisure time.

The children in this study spent the majority of the time in sitting (89.9%). This is higher than that reported by Nettlefold et al (2011) among a cohort of much younger children (8-11 years). A positive finding was the dominance of active sitting in the classroom. In other words, the children did not sit still. This finding may have implications for the health of the schoolchildren as the observed movements associated with active sitting may activate non-exercise activity thermogenesis (NEAT). NEAT is the energy expenditure that accompanies physical activities (such as fidgeting) other than volitional exercise (Levine et al, 1999). A certain threshold of energy expenditure is required for non-exercise activity thermogenesis to be activated and these thresholds are individual to each person (Levine et al, 1999). The predominance of active sitting among the children in this study indicates that this threshold of energy expenditure may be reached by many children while they are sitting in class. This has potential for further research with more objective physiological methods for measuring the energy expenditure. NEAT contributes to weight management and overall health, and as such would be positive in light of the current epidemic of global obesity in children (WHO, 2014). It could be the case that active sitting uses as much energy as activities that are traditionally classified as 'not sedentary' e.g. standing. Further investigation is required to test this hypothesis.

The German children sat more actively than the Irish children. One explanation for the difference could be the school furniture. The dynamic chairs used in the German schools promoted greater activity and therefore less static contact with the backrest. This is similar to the findings of a study on office chairs (Ellegast et al, 2012). The classroom behaviour pattern was not affected by gender. The girls moved to a greater extent than the boys when they were seated but the difference was not significant. This is in contrast to Nettlefold et al (2011) who reported that girls (74.2%) spent a significantly greater proportion of their time in sedentary behaviour than boys (71.2%). As expected, when children were attending practical subjects they engaged in significantly less sitting time than when participating in non-practical subjects. It was noted that Irish children moved around more as they completed their practical assignments. Although the German children sat to a greater extent than Irish children in practical classes, they worked while seated and therefore sat more actively than the Irish children in practical classes. A limitation of the study is that observing a child for an entire school day is labour intensive. With such a time consuming method it is difficult to obtain a large sample size, although the sample size here compares favourably with a previous similar study (Ciccarelli et al, 2011).

Conclusion

The children were seated for the majority of time in the classroom with the majority of sitting time spent sitting actively, using different sitting positions. It was noted that German children sat more actively than Irish children and there was no significant gender difference in the classroom behaviours. There was an association between classroom behaviour and class subject category. Addressing sedentary activity, which is a modifiable lifestyle behaviour, offers a public health opportunity to schools to influence activity levels.

Acknowledgements

We would like to thank all the schools and students who volunteered to participate in this research study. We would also like to acknowledge the European Commission for funding this project (Project identifier: FP7-251330-TFE).

References

Anderson, D. Field, D. Collins, P. Lorch, E. Nathan, J. (1985) Estimates of young children's time with television: a methodological comparison of parent reports with time-lapse video home observation. *Child Development;* 56: 1345-1357.

Arundell L, Ridgers ND, Veitch J,Salmon J, Hinkley T, Timerio A (2013). 5-year changes in afterschool physical activity and sedentary behaviour. American Journal of Preventative Medicine; 44(6): 605-611.

Balague, F. Troussier, B. Salminen, J. (1999) Non-specific low back pain in children and adolescents: risk factors. *European Spine Journal*; 8: 429-438.

Ciccarelli M, Straker L, Mathiassen S, Pollock C. (2011) ITKids Part 1: Children's occupations and use of information and communication technologies. *Work;* 38: 401-412.

Donnelly J, Green J, Gibson C. (2009) Physical activity across the curriculum (PAAC): a randomised controlled trial to promote physical activity and diminish overweight and obesity in elementary schoolchildren. *Preventive Medicine* 49, 336-341.

DuRant, R. Baranowski, T. Johnson, M. Thompson, W. (1994) The relationship among television watching, physical activity, and body composition of young children. *Pediatrics;* 94, 4: 449-455.

Ellegast, R. Kraft, K. Groenesteijn, L. Krause, F. Berger, H. Vink, P. (2012) Comparison of four dynamic office chairs with a conventional office chair: Impact upon muscle activation, physical activity and posture. *Applied Ergonomics*; 43, 2: 296-307.

Hamilton M, Hamilton D, Zderic T. (2007) Role of low energy expenditure and sitting in obesity, metabolic syndrome, type 2 diabetes, and cardiovascular disease. *Diabetes;* 56: 2655-2667.

Levine J, Eberhardt N, Jensen M. (1999) Role of non-exercise activity thermogenesis in resistance to fat gain in humans. *Science*; 283: 212 -214.

Matthews, C. Chen, K. Freedson, P. Buchowski, M. Beech, B. Pate, R. Troiano, R. (2008) Amount of time spent in sedentary behaviors in the United States, 2003-2004. *American Journal of Epidemiology;* 167: 875-881.

Mullen SP, Olson EA, Phillips SM, Szabo AN, Wojcicki TR, Mailey EL, Gothe NP, Fanning JT, Kramer AF, McAuley E. (2011) Measuring enjoyment of physical activity in older adults: invariance of the physical activity enjoyment scale (paces) across groups and time. *International Journal of Behavioural Activity and Physical Activity;* 8(1): 103. Open Access: 10.1186/1479-5868-8-103. PMID: 21951520

Nettlefold, L. McKay, H. Warburton, D. McGuire, K. Bredin, S. Naylor, P. (2011) The challenge of low physical activity during the school day: at recess, lunch and in physical education. *British Journal of Sports Medicine*; 45, 10: 813-819.

Owen N, Leslie E, Salmon J, Fotheringham M. (2000) Environmental Determinants of Physical Activity and Sedentary Behavior. *Exercise and Sports Science Reviews;* 28, 4: 153-158.

Ridgers N, Salmon J, Ridley K, O'Connell E, Arundell L, Timpero A (2012). Agreement between activPAL and ActiGraph for assessing children's sedentary time. International *Journal of Behavioral Nutrition and Physical Activity*; 9, 15

Rowe, P. Van der Mars, H. Schuldheisz, J. Fox S. (2004) Measuring students' physical activity levels: validating SOFIT for use with high-school students. *Journal of Teaching in Physical Education;* 23: 235-251.

Sedentary Behaviour Research Network (2012) Letter to the Editor: Standardize use of the terms "sedentary" and "sedentary behaviours". *Applied Physiology, Nutrition and Metabolism;* 37: 540-542.

Sims S, Larson J, Lamonte M, Michael Y, Martin L, Johnson K, Sarto G, Stefanick M. (2012) Physical activity and body mass: changes in younger vs. older postmenopausal women. *Medicine and Science in Sport and Exercise;* 44, 1: 89-97.

Smith A, Lockwood C, Moon J, Kendall K, Fukuda D, Tobkin S, Cramer J, Stout J. (2010) Physiological effects of caffeine, epigallocatechin-3-gallate, and exercise in overweight and obese women. *Applied Physiology, Nutrition and Metabolism;* 35, 5: 607-616.

Tremblay M, Colley R, Saunders T, Healy G, Owen N. (2010) Physiological and health implications of a sedentary lifestyle. *Applied Physiology Nutrition and Metabolism;* 35, 6: 725-740.

Tremblay M, LeBlanc A, Janssen I, Kho ME, Hicks A, Murumets K, Colley R, Duggan M. (2011A) Canadian sedentary behaviour guidelines for children and youth. *Applied Physiology and Nutrition Metabolism;* 36: 59-64.

Tremblay, M. LeBlanc, A. Kho, M. Saunders, T. Larouche, R. Colley, R. Goldfield, G. Gorber, S. (2011B) Systematic review of sedentary behaviour and indicators in school-aged children and youth*, International Journal of Behavioural Nutrition and Physical Activity*; http://www.ijbnpa.org/content/8/1/98

Treuth, M. Schmitz, K. Catellier, D. McMurray, G. Murray, D. Almeida, M. Going, S. Norman, J. Pate, R. (2004) Defining accelerometer threshold for activity intensities in adolescent girls. *Medicine and Science in Sport and Exercise;* 36: 1259-1266.

Treuth, M. Catellier, D. Schmitz, K. Pate, R. Elder, J. McMurray, R. Blew, R. Yang, S. Webber, L. (2007) Weekend and weekday patterns of physical activity in overweight and normal-weight adolescent girls. *Obesity*; 15, 7, 1782-1788.

Van Uffelen, J. Wong, J. Chau, J. (2010) Occupational sitting and health risks: a systematic review. *American Journal of Preventative Medicine*; 39, 4, 379-388.

Vereecken, C. Todd, J. Roberts, C. Mulvihill, C. Mae, L. (2006) Television viewing behaviour and associations with food habits in different countries*. Public Health and Nutrition;* 9: 244-250.

Watson, K. Papageorgiou, A. Jones, G. Taylor, S. Symmons, D. Silman, A Macfarlane, G. (2002) Low back pain in schoolchildren: occurrence and characteristics. *Pain*; 97, 87-92.

World Health Organisation (2014). Global strategy on diet, physical activity and health. http://www.who.int/dietphysicalactivity/childhood_why/en/: Accessed on 3rd July 2014.

HUMAN FACTORS TRAINING FOR HOSPITAL DOCTORS IN IRELAND

Eva Doherty D.Clin Psych., Dara O'Keeffe MRCS

National Surgical Training Centre, Royal College of Surgeons in Ireland, 123, St. Stephens Green, Dublin 2

Abstract

Human factors concerns the interactions between people and technical components in complex systems. In medicine, these factors are recognised as potential sources of adverse events in hospitals and it is estimated that one in ten hospital admissions will involve an adverse event. The evidence is increasing that team-based initiatives and training in human factors is associated with significant reductions in adverse events and improvements in patient outcomes. The Human Factors and Patient Safety training programme for hospital doctors at the Royal College of Surgeons in Ireland is described. The doctors' evaluations indicate a high level of satisfaction with the training provided and assessments conducted after two years of training indicate improvements in a range of communication skills with patients, relatives and colleagues.

Introduction

Patient safety is a topic of increasing concern in Irish hospitals and the Human Factors model provides a framework which helps to both understand and prevent human error. Based on international evidence it is estimated that one in ten hospital admissions will involve an adverse event (Bognor, 1994). A significant proportion of adverse clinical events are linked to surgery which is a high risk area and non-technical errors are thought to be much more common than technical errors. (Catchpole, 2010). Communication skills deficits have been found to be associated with an increased risk of complaints to medical regulatory authorities (Tamblyn et al), but the communication skills of residents have been shown to improve following a series of six communication training sessions (Hochberg et al.,2010).

Various forms of human factors instruction have been shown to have a positive effect on physician performance and on patient safety. Crew resource management techniques are being used to develop training programmes for medical personnel incorporating decision-making, communication, leadership training. (Hughes et al., 2014). Attendance at leadership training is associated with enhanced leadership competencies as assessed by the likelihood of adopting a leadership role on hospital and national committees (Day et al., 2010), and team training has been shown to significantly reduce surgical mortality (Neily et al.,2010).

The Human Factors programme in RCSI initially focused on teaching human factors to surgical trainees but has expanded to include other specialties and disciplines including Emergency Medicine.

Description of Human Factors Training

The Royal College of Surgeons in Ireland (RCSI) established a training programme in Human Factors and Patient Safety ten years ago. The programme is a mandatory component of training at RCSI and trainees in multiple disciplines attend for three full days instruction per year. This results in 21 days of instruction in Human Factors and Patient Safety for each trainee over the course of their training programme.

Course content and participants

At RCSI, trainees in various specialties receive instruction on the Human Factors and Patient Safety programme. The course content of the Human Factors and Patient Safety programme is described in Table 1. Teaching is learner-focused and active rather than passive. Didactic teaching is kept to a minimum and learning outcomes are addressed through classroom activities as much as possible. Tasks are set to groups of trainees to facilitate their direct experience of team skills and leadership abilities. Communication skills for challenging situations are taught through the use of role play exercises both with each other and with professional actors. High fidelity simulation scenarios are employed for enhanced engagement in suitable modules. At the more senior level, the programme takes account of the advanced learning requirements of the trainees. Experts from a variety of backgrounds support the teaching e.g., pilots, indemnity representatives, lawyers, medical council officers and clinicians and psychologists are invited to co-facilitate specific teaching sessions.

Table 1. Human Factors and Patient Safety programme: Course content.

Year of training	Module 1	Module 2	Module 3
Year 1	Error and safety in hospital practice	Talking to patients and relatives	Crisis management
Year 2	Teamwork	Conflict resolution	Decision-making and disclosure of error
Year 3	Personality and behaviour	Emotional intelligence and stress management	Leading teams
Year 4	Principles of effective handover	Critical Incident Analysis	Crisis management simulation I

Year 5	Ethical problems in hospitals	Breaking bad news	Change programmes in healthcare delivery
Year 6	Disclosure of error	Crisis management simulation	Motivational Interviewing skills for challenging situations
Year 7	Leadership 1	Leadership 2	Leadership 3

Assessment

Trainee satisfaction

We collected post class participant feedback from a total of 420 trainees over a period of one academic year (2013-2014). The data represented in Figure (1) below was measured on a Likert scale of 1-5, with 5 being the highest satisfaction rating. The feedback indicates that the mean scores were between 4 and 5 for each topic.

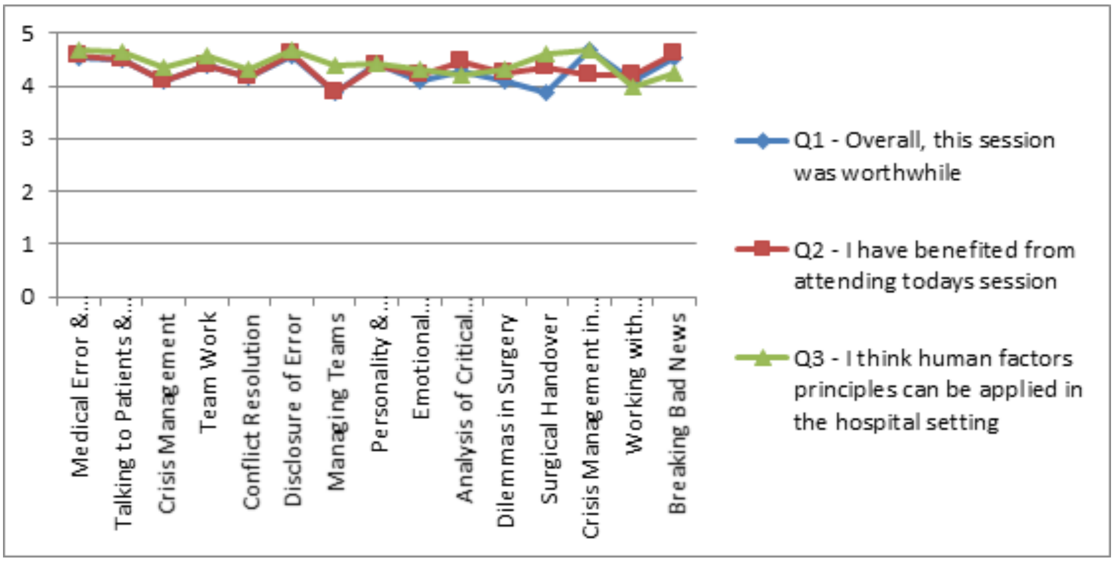


Figure 1. Trainee satisfaction with Human Factors training topics at RCSI (N=420).

Improvement in communication skills

We also assessed the improvement in communication skills in a sample of 50 trainees over the course of two years training. The examinations consisted of OSCE stations comprised of simulated scenarios using professional actors which reflected topics covered during the training programme, specifically obtaining consent, colleague conflict and disclosing medical error. Each station was scored using a standardised marking sheet. Consultant surgeons of various specialties acted as examiners. The examiners received training prior to the commencement of the examination. This training was run by the clinical psychologist and lecturer in surgical education who established the training programme. Trainees were assessed in the scenarios early on the first day of training, before they received any instruction and again at the end of their second year of training. They had therefore attended for a total of six sessions.

Statistical analysis of the mean scores of the assessments of each scenario indicates that the communication skills demonstrated by the trainees in year two were significantly better than those demonstrated by these trainees in year one. This was the case for all scenarios- consent (t (49) = -5.39; $P < 0.001$) (two tailed), colleague conflict (t (48) = -8.63; $P < 0.001$) (two tailed) and disclosure of error (t (49) = -7.56; $P < 0.001$) (two tailed). Furthermore upon calculating the effect size, it was found that the year trainees were in had a large effect on the scores obtained for all scenarios. Specifically it was found that the year trainees were in accounted for 23% of the variance in the scores obtained from the consent OSCE (eta squared = .232), 44% of the variance in the scores obtained from the colleague conflict OSCE (eta squared = .437) and 37% of the variance in the scores obtained from the disclosure of error OSCE (eta squared = .368). These results indicate that significant improvements in communication skills occurred over the course of human factors training. As this was not a controlled trial, it cannot be assumed that this was directly the result of attending human factors training however the finding does contrast with international concern that empathy and communication skills deteriorate over time spent in medical training (Mikesell, 2013, Batt-Rawden et al., 2013, Hojat et al., 2009).

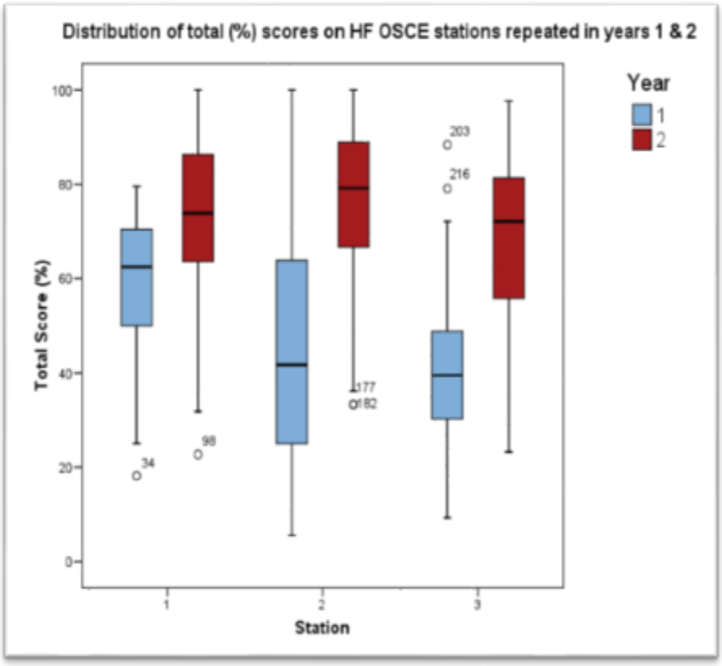


Figure 2. Improvement in the communication skills of N=50 trainees attending a Human Factors and Patient Safety training programme over two years.

Conclusions

A new era of acceptance that human factors are essential for the development of high performing physicians and the maintenance of excellent patient safety is dawning. The Irish

Medical Council has outlined eight domains of good professional practice, and six of these eight domains describe skills or attitudes that are within the realm of human factors. As outlined in our introduction, it has been demonstrated that these skills can be taught and will impact on the quality of healthcare delivery. The Royal College of Surgeons in Ireland is a world leader in the inclusion of mandatory training of these skills for non-consultant hospital doctors. By formally teaching a Human Factors and Patient Safety programme to non-consultant hospital doctors, we aim to contribute to a profession-wide culture change that values these skills.

References

Batt-Rawden, S.A., Chisolm, M.S., Anton, B. 2013,Teaching empathy to medical students: An updated systematic review, *Academic Medicine*,**88(8),**1171-1177

Bogner, M. 1994, *Human Error in Medicine*. Hollsdale, NJ: LEA

Catchpole, K. 2010, Errors in the Operating Theatre-How to Spot and Stop Them. Jounalof Health Services Research and Policy,**15(Suppl 1),**48-51

Day, C.S., Tabrizi, S., Kramer, J., Yule, A.C,, Ahn, B.S. 2010, Effectiveness of the AAOS Leadership Fellows Program for Orthopaedic Surgeons, *Journal of Bone and Joint Surgery*, **92**,2700-2708

Hochberg, M.S., Kalet, A., Zabar, S., Kachur, E., Gillespie, C., Berman, R.S. 2010, Can Professionalism be Taught ? Encouraging Evidence. *American Journal of Surgery*, **199**,86-93

Hojat, M.,Vergere,M.J.,Mavell,K.,2009, The devil is in the third year: A longitudinal study of the erosion of empathy in medical school, *Academic Medicine,***84**(9),1182-1191.

Hughes, K.M., Benenson, R., Krichten, A.E., Clancy, K.D., Ryan, J.P., Hammond, C. 2014, A Crew Resource Management Program Tailored to Trauma resuscitation Improves Team Behavior and Communication. *Journal of the American College of Surgeons*, **219**,545-551

Mikesall,L.2013,Medicinalrelationships:Caringconversation,*MedicalEducation,***47(5),**443-452

Neily, J., Mills P.D., Young-Xu Y..Carney, B.T.,West, P., Berger, D.H., Mazzia, L.M., Paull, D.E., Bagian, J.P. 2010, Association Between Implementation of a Medical Team Training Program and Surgical Mortality, *Journal of the American Medical Association*, **304(15),**1693-17000

Tamblyn, R., Abrahamowicz, M., Dauphinee, D., Wenghofer, E., Jacques, A., Klass,D., Smee, S., Blackmore, D., Winslade, N., Girard, N., Du Berger, R., Bartmann, I., Buckeridge, D.L., Hanley, J.A. 2007. Physician Scores on a National Clinical Skills Examination as Predictors of Complaints to Medical Regulatory Authorities, *Journal of the American Medical Association*, **298(9),** 993-1001

AN INVESTIGATION OF THE RELATIONSHIP OF PERCEIVED WORKLOAD AND 'LEVELS OF AUTOMATION' OF RADIATION ONCOLOGIST AT THE 'TREATMENT DELIVERY' STAGE OF LOW-DOSE RATE BRACHYTHERAPY

M. Galičič, E. F. Fallon, W. Van der Putten

Irish Centre for Patient Safety,
National University of Ireland, Galway,
Galway

Abstract

Low-Dose Rate (LDR) brachytherapy is a successful method for treatment of prostate cancer with a high patient survival rate. The perceived workload of Radiation Oncologist was measured using the NASA-TLX technique in a public hospital in Ireland. A task analysis of the Radiation Oncologist's operations at the "Treatment delivery" stage was developed using Hierarchical Task Analysis (HTA). A "Modified Levels of Automation" model was applied to the main operations. The results have shown that the average levels of perceived workload were low, and that the applied 'levels of automation' were low and medium. The operations with the highest perceived workload median were "Peripheral seeding and planning", "Needle implantation" and "Central seeding and planning". A low correlation was found between the reported workload levels and levels of automation used in the present system.

Introduction

Low-Dose Rate brachytherapy

Low-Dose Rate (LDR) brachytherapy is a one-off treatment procedure where small radioactive seeds (usually Iodine-125; or I-125) are permanently implanted into the patient's prostate gland. "Brachy" means "short" in Greek, and refers to the Radiation Oncologist implanting the seeds at a "short distance" while the patient is anaesthetised. The seeds, stored in a seed cartridge, are carefully implanted via an applicator through the needles. The objective of the treatment procedure is to reach an even radiation dose distribution throughout the prostate gland. The radiation dose is determined by specialised software which also creates 2D & 3D models of the implanted seeds. The radioactivity of the implanted seeds destroys cancerous cells. The half-life of the implanted I-125 seeds is approximately 60 days. The success of this procedure depends on a highly trained team of professionals and the use of advanced automated equipment.

A generic flowchart of LDR brachytherapy is presented in Figure 1. In general, after the patient is diagnosed with cancer, they are referred to the Radiation Oncology department where imaging for radiotherapy planning is performed, usually by using Computerised Tomography (CT) or Magnetic Resonance Imaging (MRI). Treatment planning includes a pre-treatment review of the images and verification of the procedure, mainly by the Radiation Oncologist and

Medical Physics team. Treatment delivery takes place at the scheduled date. Straight after the treatment delivery, the patient is taken to a recovery room. Then another image of the patient's prostate gland is performed before they are discharged.

LDR prostate brachytherapy has a relatively high survival rate. According to (Kibel et al. 2012), adjusted 10-year survival rate of the patients who had LDR brachytherapy was 81.7%.

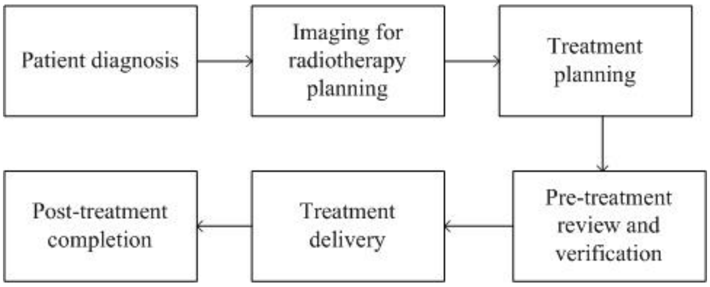


Figure 1. A generic flowchart of LDR prostate brachytherapy

Workload

The workload can broadly be defined as "a required demand for the human" or in terms of human factors as "the ratio of the available resources and the required resources during the task" (Wise et al. 2010). It directly influences human performance. Workload can be measured objectively (e.g. heart rate, brain wave, etc.) or subjectively (assessment techniques and scaling tools) based on the perceived experience of the operator reflecting the nature of the operation, working environment, and pace of work. The workload should be "optimal" and should achieve "optimal" performance (see ISO 10075). If workload is too low, operators may get bored, fatigued, unengaged, their attention can decrease, they can process information incorrectly, etc. If workload is too high, the operators may get too stressed and unable to cope with the pace of work and their decision making may be compromised under pressure. Both, low and high workload may create a potential for errors, which can make the treatment procedure unsafe, and it can expose the patient and team of professionals present at the treatment procedure to risk. The optimal workload is somewhere between low and high and can have a positive effect on job satisfaction, job performance, organisational safety and staffing levels. NASA-TLX, a subjective workload technique was used in this research to investigate the levels of workload at "Treatment delivery" stage of LDR brachytherapy treatment.

The 'Modified Levels of Automation' Model for Brachytherapy

'Levels of Automation' (LOA) model utilises the application of function allocation between human operators and machines, i.e. automation. LOA models by Parasuraman et al. (2000) and Sheridan (1997) represent ten levels of automation applying the function allocation between human operators and automated equipment, between manual and fully automated. The application of these LOA models lacked clarity at level 10, full automation. This level had to be modified to fit the LDR brachytherapy treatment. A 'Modified Levels of Automation' (MLOA) model splits the tenth level of "full automation" into two sub-classes: a new class (a), when the human operators have the ability to veto and to stop the system or operations, and the existing class (b) originally presented by Parasuraman et al. (2000) and Sheridan (1997), when this ability is not available. However, the four categories of perception remain the same: 1. Information Acquisition; 2. Information Analysis; 3. Decision Selection; and 4. Action Implementation.

Methods

Hierarchical Task Analysis

The Hierarchical Task Analysis (HTA) was used to represent the operations and sub-operations at the "Treatment delivery" stage of LDR brachytherapy. "Operations" are actions performed by human operator in order to achieve the goal. HTA is a well known and commonly used human factors method for task analysis of work processes (Stanton et al. 2013). It represents operation scenarios in a hierarchical order, including the components, the order of the steps, goals, sub-goals, operations and plans. It can be presented as a diagram or as a table. HTA has been used extensively in healthcare (Lane et al. 2006; Phipps et al. 2008).

NASA-TLX

The National Aeronautics and Space Administration (NASA) Task Load Index (TLX) was developed by (Hart & Staveland 1988). It is a well established multi-dimensional human factors method for measuring perceived workload. It is often referred to as "the most commonly used subjective workload assessment technique" (Stanton et al. 2013). It has been applied to various environments, from nuclear industry, aviation, military domains, automotive, to healthcare (Stanton et al. 2013). A paper-based NASA-TLX technique was used, using so called "TLX worksheets" which were administered straight after the operations were completed. A single TLX worksheet includes assessment of six dimensions (also referred to as "TLX dimensions"): Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (PE), Effort (EF), and Frustration (FR). For a detailed description of TLX dimensions see Table 1. The dimensions are measured on a scale from 0 to 100 for each dimension and are rounded to the nearest 0.5. The assessment includes a pair-wise comparison between each of the six dimensions. This is to weight the relevance of each dimension. The overall workload score is calculated by combining the results from each dimension and pair-wise comparison weights, and dividing them by 15 (the total number of paired comparisons). The results reflect low, moderate or high levels of perceived workload of the assessed operation.

A workload rating scale (Galičič et al. 2013) presented in Table 2 was used to categorise the TLX results into low, moderate or high. The cut-off point for high workload was 55 or greater, as presented in similar research by (Mosaly et al. 2011; Mazur et al. 2012).

Table 1. NASA-TLX dimensions

TLX dimension	Abbreviation	Endpoint	Description

Mental Demand	MD	Low / High	How mentally demanding was the task? E.g. thinking, deciding, calculating, remembering, searching for, etc.
Physical Demand	PD	Low / High	How physically demanding was the task? E.g. Pushing, pulling, turning, controlling, gripping, adjusting, etc.
Temporal Demand	TD	Low / High	How hurried or rushed was the pace of the task? E.g. Fast, medium or slow pace
Performance	PE	Good / Poor	How successful were you in accomplishing what you were asked to do?
Effort	EF	Low / High	How hard did you have to work to accomplish your level of performance?
Frustration	FR	Low / High	How insecure, discouraged, irritated, stressed, and annoyed were you?

Table 2. Workload rating scale (Galičič et al. 2013)

Workload level	Workload score level
Low	< 35
Moderate	35 - 55
High	> 55

The 'Modified Levels of Automation' (MLOA) Model for Brachytherapy

The 'Modified Level of Automation' (MLOA) model expands the 10th "full automation" stage of the 'Level of automation' model from Parasuraman (2000) and Sheridan (1997) into two sub-classes: (a) human operator having an option to veto and stop the machine processes (10a), and (b) human operator keeping the non veto option, leaving all decisions to the machine (10b) (see Table 3). These two sub-classes will be crucial when differentiating the roles between man and machine in more automated systems, which will be in use in the future. The four levels of perception however, remain unchanged: Information acquisition, Information analysis, Decision selection, and Action implementation.

Table 3. Expanded 10th 'Modified Level of Automation' with levels '10a' and '10b'.

LOA	Information Acquisition	Information Analysis	Decision selection	Action Implementation
10a	The computer executes everything, acts autonomously, human can veto at any time	The computer does all cognitive functions, acts autonomously, human can veto at any time	The computer decides everything, acts autonomously, human can veto at any time	The computer executes everything, acts autonomously, human can veto at any time
10b	The computer executes everything, acts autonomously, without the human	The computer does all cognitive functions, acts autonomously, ignoring the human	The computer decides everything, acts autonomously, ignoring the human	The computer executes everything, acts autonomously

Automation deployment and description of the LDR operations

In this research, the function allocation was between Radiation Oncologist and Trans-Rectal Ultra-Sound (TRUS). TRUS gives Radiation Oncologist a visual feedback, based on which the decisions are made on the number and location of the implanted radioactive seeds. TRUS is also connected to the computer with specialised software, operated by Medical Physicist. The software calculates the isodose of the implanted seeds and creates 2D & 3D models of the implanted seeds which are then reviewed. The software can also serve as an isodose simulation of the pre-implanted seeds to help with the decision making on the number and the location of the implanted seeds. The LOA were applied to each operation of the Radiation Oncologist "Treatment delivery" stage of the LDR brachytherapy treatment, and are presented in Table 7.

Procedure

The following steps were taken in this research:

- Development of Hierarchical Task Analysis.
- Workload measurement using NASA-TLX technique.
- Application of 'Modified Levels of Automation'.
- Correlation of workload and 'Modified Levels of Automation'.
- Presenting the summary of research.

Results

Hierarchical Task Analysis

The Hierarchical Task Analysis results are presented in a tabular format. The "Treatment delivery" stage had 8 operations and 37 sub-operations, as presented in Appendix 1.

Workload

Fifteen LDR brachytherapy treatment cases were assessed for perceived workload by using the NASA-TLX technique. Each case assessed eight Radiation Oncologist's operations. One hundred and twenty TLX worksheets were administered altogether (15x8) to one Radiation Oncologist at the "Treatment delivery" stage of LDR brachytherapy.

The following results were obtained:

1. The overall workload scores for each administered TLX worksheet are presented in Table 4 and Figure 2.
2. The percentage proportion of six TLX dimensions is presented in Table 5.
3. The overall mean proportion of six TLX dimensions is presented in Figure 3.

The three highest overall workload medians for each operation were classified as "low" workload:

- 32.66 at "Peripheral seeding and planning" operation;

• 30.00 at “Needle implantation” operation;

• 27.00 at “Central seeding and planning” operation.

Despite the workload classified as “low” at three operations, four out of eight operations reported “high” workload levels above 55 as a maximum score (see Table 4). These operations were: "TRUS placement" operation, "Needle implantation" operation, “Virtual prostate and urethra positioning” and "Central seeding and planning" operation.

The three dimensions with the highest proportions in each operation presented in Table 5 were:

• TRUS placement: MD (47.2%), EF (35.8%) and TD (9.7%)

• Measure volume: MD (28.6%), EF (37.6%) and TD (19.1)

• Needle implantation: EF (42.9%), MD (40.2%), and FR (5.4%)

• Volume post needle: MD (48.5%). MD (25.7%) and TD (13.6%)

• Virtual prostate and urethra positioning: MD (38.5%), FR (24.5%) and TD (13.7%)

• Peripheral seeding and planning: MD (47.5%), EF (31.0%) and TD (8.0%)

• Central seeding and planning: MD (57.4%), EF (26.9%), and MD (7.7%)

• Volume post implant: EF (32.4%), MD (25.9%) and TD (21.1%).

The three dimensions with the highest overall mean were MD (41.7%), EF (30.5%) and TD (12.1%) for the following operations: “Peripheral seeding and planning”, “Needle implantation”, and “Central seeding and planning”. These three operations are presented hierarchically from the highest to the lowest in Table six.

The overall mean of the TLX dimensions presented in Table 5 and the mean of the TLX dimensions presented in Table 6 were compared. The mean of the TLX dimensions from Table 6 reported higher MD by 6.7%, higher EF by 3.1% and higher PD by 0.7%; lower TD by 6.5%, lower FR by 4.1%, and lower EF by 3.1%.

‘Modified Levels of Automation’

The ‘MLOA’ tool was applied to the eight operations carried out by the Radiation Oncologist. The majority of operations were found to have low, and some had medium level of automation. The medium level of automation was at the “Information Acquisition” stage. For details, see Table 7 and Figure 4. The results correspond to the levels of automation in the current LDR brachytherapy system (Fallon et al. 2015).

The relationship of the perceived workload and ‘Modified Levels of Automation’
The values of the perceived workload were adjusted by one decimal place to a 10-point scale. The mean values of LOA were calculated for each operation. Pearson’s Correlation coefficient was used to correlate the perceived workload levels and levels of automation. Pearson’s Correlation coefficient was 0.142, which suggests a low correlation between two variables.

Table 4. Overall TLX workload scores

	Operation	N	Min.	Max.	Mean	Median	SD	Workload level
1	TRUS placement	15	2	59	21.69	15.33	16.32	Low
2	Measure volume	15	0	22	10.83	9.00	6.82	Low
3	Needle implantation	15	16	63	36.37	30.00	15.39	Low
4	Volume post needle	15	3	33	13.60	12.33	7.85	Low
5	Virtual prostate and urethra positioning	15	0	64	9.99	7.33	15.56	Low
6	Peripheral seeding and planning	15	0	48	30.40	32.66	15.36	Low
7	Central seeding and planning	15	6	60	26.81	27.00	15.50	Low
8	Volume post implant	15	0	28	8.49	6.00	8.25	Low

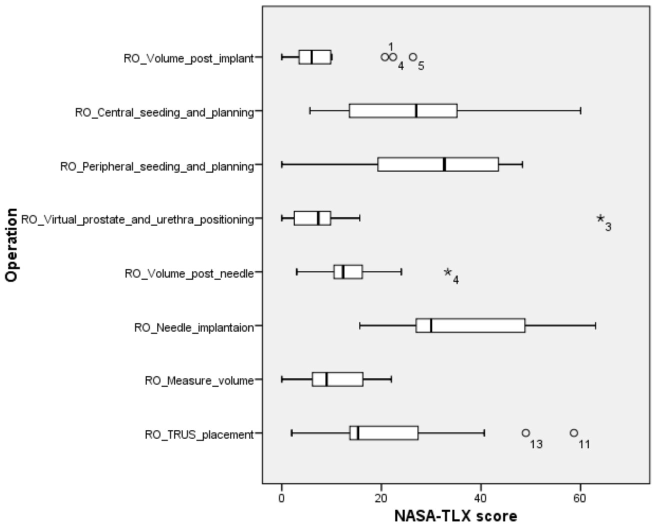


Figure 2. Overall TLX workload scores for eight sub- operations of Radiation Oncologist (RO)

Table 5. Percentage proportion of six TLX dimensions

Operation number	MD	PD	TD	PE	EF	FR
1	47.2	2.7	9.7	1.4	35.8	3.3
2	28.6	3.0	19.1	4.6	37.6	7
3	40.2	3.7	3.8	4.0	42.9	5.4
4	48.5	2.9	13.6	3.7	25.7	5.6
5	38.5	1.3	13.7	10.2	11.8	24.5
6	47.5	4.8	8.0	3.3	31.0	5.4
7	57.4	2.7	7.7	3.5	26.9	1.8
8	25.9	3.1	21.1	4.4	32.4	13.0
Overall mean	41.7	3.0	12.1	4.4	30.5	8.3

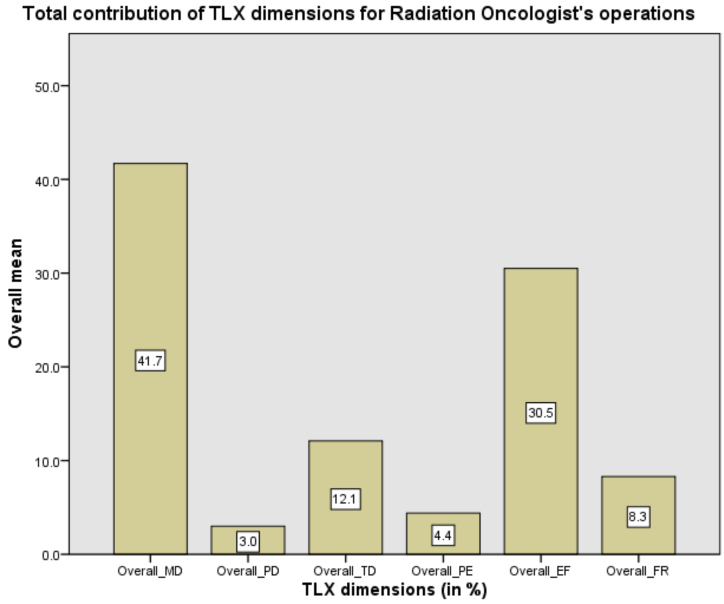


Figure 3. The overall mean proportion of six TLX dimensions

Table 6. Percentage values of the TLX dimensions (rounded to the nearest decibel place)

The highest workload level rank	Operation		Median	Workload score	MD	PD	TD	PE	EF	FR
	Number	Description								
1	6	Peripheral seeding and planning	32.66	Low	47.5	4.8	8.0	3.3	31.0	5.4
2	3	Needle implantation	30.00	Low	40.2	3.7	3.8	4.0	42.9	5.4
3	7	Central seeding and planning	27.00	Low	57.4	2.7	7.7	3.5	26.9	1.8
				Mean	48.4	3.7	6.5	3.6	33.6	4.2

Table 7. Application of the 'Modified Levels of Automation' model of Radiation Oncologist's operations

Operation		Modified Level of Automation			
		Information Acquisition	Information Analysis	Decision selection	Action Implementation
1	TRUS placement	1	1	1	1
2	Measure volume	2	2	2	2
3	Needle implantation	2	2	2	2
4	Volume post needle	5	2	2	2
5	Virtual prostate and urethra	5	2	2	2

	positioning				
6	Peripheral seeding and planning	5	2	2	2
7	Central seeding and planning	5	2	2	2
8	Volume post implant	2	2	2	2

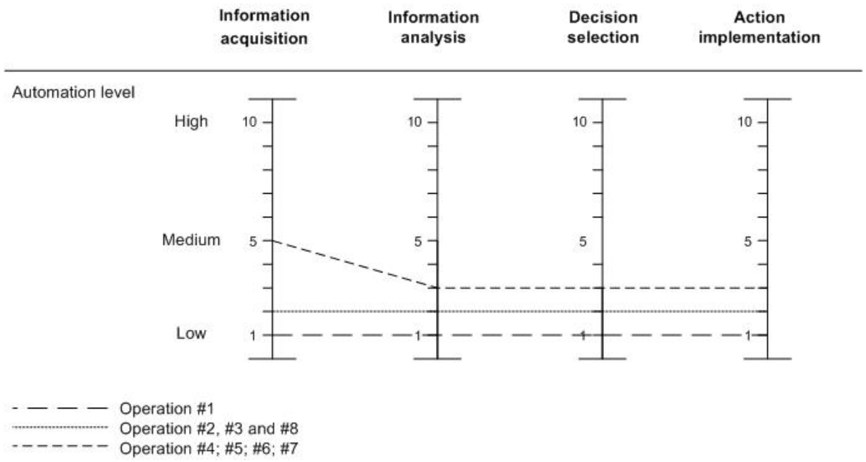


Figure 4. A graphical representation of the 'MLOA' for Radiation Oncologist's operations

Discussion

Hierarchical Task Analysis

Eight operations were noted based on the observation of over twenty cases of LDR brachytherapy by the researchers and with the help from the staff working on those cases prior to the data collection. Each operation represents a set of sub-operations which are performed in a systematic manner in order to keep the most optimal performance of the LDR brachytherapy treatment process. Eight operation "Volume post implant", was an experimental operation, and is not crucial to the LDR brachytherapy treatment process.

It is important to note that HTA was kept as short and concise as possible. It is also important to note that Radiation Oncologist's operations and subsequent operations are not standalone operations. They are often performed in cooperation with the other team members during this procedure: Nurse, Anaesthetist and Medical Physics team. Their operations sometimes overlap, e.g. Nurse helping out the Radiation Oncologist with the preparation of the seed cartridges, the applicator, etc., Medical Physics team working on a 2D & 3D representation of the isodose and quality assurance of the implanted seeds, and Anaesthetist monitoring the patient while they are anaesthetised.

Workload

The operations with the highest perceived workload were classified as “low” based on their median. Three operations with the highest median are presented in Table 6. When compared to the values in Table 5, the following differences were noted:

MD was higher because the operations required higher mental and perceptual activity. EF was higher because the operations were more mentally and physically demanding than in other cases. A slight rise in PD has shown almost no difference when compared with the overall mean values. However, there was a decrease of TD indicating the pace of the operations was lower. FR was lower, indicating the lower levels of stress. PE was also lower, indicating the performance level was not affected. The results also show that these three operations are the most mentally demanding, and require the most effort during the LDR brachtherapy treatment case. Therefore we can say that these operations require the most focus, training and high levels of skills to be performed, and that they are crucial in terms of safe treatment delivery. Moreover, both, the “Peripheral and “Central seeding and planning” operations are crucial in terms of the correct contouring, that is labelling the prostate in 0.5 increments on the ultrasound, determining the prostate volume, the shape of the prostate gland, and any possible interference, e.g. pubic arch and urethra. The objective of the LDR brachytherapy is to insert the needles and to implant the seeds correctly in order to eradicate cancerous cells. They can only be eradicated with the correct isodosis of the prostate gland. Both of these operations compliment the “Needle implantation” operation, which, when needle is inserted into the right position, determines the position of the implanted I-125 seeds. Despite the perceived workload results being reported as “low”, they depend on the perceived workload of the individual person assessed. The results may reflect high skills and knowledge and years of training and experience Radiation Oncologist has to undergo before they are qualified to deliver such treatment. This means that a newly trained Radiation Oncologist would most possibly rate these operations higher. It would be interesting to investigate workload at other hospitals in Ireland and abroad, e.g. in the UK or in the USA.

It also has to be noted that this paper only looked at the median values of the overall TLX scores. It did not look into the individual LDR brachytherapy treatment cases which reported the high perceived workload values of 63, 60 and 48. These cases would need to be researched individually, which was not the objective of this paper.

Another argument is that despite the perceived workload levels being reported as “low”, it is not possible to determine when these levels are still accepted as “low”, and when they are “too low”. Levels reported as “too low” would reflect the negative workload effects. However, the researchers did not note any signs of negative workload effects. The treatment delivery was well paced and skilfully performed, with individuals working well in a team.

‘Modified Levels of Automation’

Interestingly, “Image Acquisition” had the highest ‘level of automation’ score of 5 at “Volume post needle”, “Virtual prostate and urethra positioning”, “Peripheral seeding and planning”, and “Central seeding and planning” operations. These operations involve Radiation Oncologist contouring the patient’s prostate on the ultrasound screen to obtain prostate volume. The ultrasound’s output can be followed in real time on the computer, connected to the ultrasound. The computer offers a few options which Radiation Oncologist approves before they are executed, e.g. 2D and 3D modelling, total dose distribution, etc.

The correlation of perceived workload and ‘Modified Levels of Automation’

The TLX data reports low workload levels, and interestingly, has low correlation with the ‘MLOA’, which reports low to medium levels of automation in the present technological system. Medium levels of automation were reported at “Peripheral seeding and planning” and “Central seeding and planning” operations. During these two operations the Radiation Oncologist acquires most of the data from the ultrasound, and based on this makes decisions on where and how many needles and I-125 seeds need to be implanted. There may be stronger correlation between workload and automation in the future, e.g. increased workload with increased levels of automation. Although based on the data collected in this paper, it is yet too early to predict such correlation.

It is important to note that this data set only reflects Radiation Oncologist’s operations at one hospital site. It would be interesting to research Radiation Oncologist’s levels of automation at other hospitals and to compare lager data sets from other hospital sites. It would also be interesting to relate the existing data with the future levels of automation, when automation levels are higher at other stages as well as other operations.

Conclusions

The TLX data reports low perceived workload levels of the operations performed by Radiation Oncologist at the “Treatment delivery” stage of LDR brachytherapy treatment. Three levels with the highest workload scores were “Peripheral seeding and planning”, “Needle implantation” and “Central seeding and planning”, which are crucial operations and could potentially compromise the safety of the treatment delivery. Interestingly, workload levels had low correlation with the ‘MLOA’ data, which predominantly reported low to medium levels of automation. However, the data set in this paper only reflects Radiation Oncologist’s operations at one hospital site. It would be interesting to research Radiation Oncologist’s workload levels and the LOA at other hospitals, and to compare more data altogether. It would also be interesting to relate the existing data with the future LOA, when automation levels are higher at these operations.

Acknowledgements

The authors would like to acknowledge the Radiotherapy staff at the participating hospital and to Health Research Board (HRB) of Ireland who provided financing for this research.

References

Fallon, E.F, Galičič, M., Chadwick, L., van der Putten, W. 2015, *An Analysis of the Impact of Trends of Automation on Human Error Potential in Brachytherapy*, Proceedings to the 6th International Conference on Applied Human Factors and Ergonomics – 4th International Conference on Human Factors and Ergonomics in Healthcare, 26-30 July, Las Vegas, Nevada, United States.

Galičič, M., Fallon E.F., van der Putten, W., Sands G., 2013, *Applying NASA-TLX to Prostate Seeds Brachytherapy*. Patient & Healthcare Provider Safety, London, November 2013.

Hart, S. G., & Staveland, L. E., 1988, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139-183.

ISO 10075:1991 - Ergonomic principles related to mental workload - General terms and definitions. Accessed on 15 April 2015 from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=18045

Kibel, A. S., Ciezki, J. P., Klein, E. A., Reddy, C. A., Lubahn, J. D., Haslag-Minoff, J., Deasy, J.O., Michalski, J.F., Kallogjen, D.K., Piccirillo J.F., Rabah D.M., Yu, C., Kattan M.W. & Stephenson, A. J., 2012, *Survival among men with clinically localized prostate cancer treated with radical prostatectomy or radiation therapy in the prostate specific antigen era*. The Journal of urology, 187(4), 1259-1265.

Lane, R., Stanton, N. A., & Harrison, D. (2006). Applying hierarchical task analysis to medication administration errors. Applied ergonomics, 37(5), 669-679.

Mazur, L. M., Mosaly, P. R., Jackson, M., Chang, S. X., Burkhardt, K. D., Adams, R. D., Jones, E. L., Hoyle, L., XU, J. & Rockwell, J., 2012, Quantitative assessment of workload and stressors in clinical radiation oncology. *International Journal of Radiation Oncology* Biology* Physics,* 83**,** e571-e576.

Mosaly, P. R., Mazur, L. M., Jackson, M., Chang, S. X., Burkhardt, K. D., Jones, E. L., Xu, J., Rockwell, J. & Marks, L. B., 2011, Empirical Evaluation of Workload of the Radiation Oncology Physicist During Radiation Treatment Planning and Delivery. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 55, No. 1, pp. 753-757). SAGE Publications.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D., 2000, A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *30*(3), 286-297.

Phipps, D., Meakin, G. H., Beatty, P. C., Nsoedo, C., & Parker, D. (2008). Human factors in anaesthetic practice: insights from a task analysis. British journal of anaesthesia, 100(3), 333-343.

Sheridan, T. B., 1997, Function allocation: algorithm, alchemy or apostasy? In E.F. Fallon, L. Bannon, J. McCarthy (Eds.), *Revisiting the allocation of function issue - New perspectives Vol. 1*, Proceedings of the First International Conference on Allocation of Functions, Louisville: IEA Press.

Stanton, N. A., Salmon P. M., Rafferty L.A. Walker, G. H., Baber C. & Jenkins, D.P., 2013, *Human factors methods: A practical guide for engineering and design*. Ashgate Publishing, Ltd.
Wise, J. A., Hopkin, V. D., & Garland, D. J. (Eds.), 2010, *Handbook of aviation human factors, 2nd edition*. CRC Press.

Appendix 1

Hierarchical Task Analysis

Operator: Radiation oncologist			
Main task: Treatment delivery			
Plan: Do all in one sequence; do optional if necessary			
Operation	**Operation Description**	**Sub-operation number**	**Detailed sub-operation description**
1	TRUS placement	1.1	Wash and disinfect hands & put on cap, shoe covers, gloves, surgical gown and mask
		1.2	Set up the TRUS
		1.3	Set up the stepper
		1.4	Set up the disposable needle template on the stepper
		1.5	Prepare the TRUS (apply the condom and the gel)
		1.6	Place TRUS probe into the patient
		1.7	Acquire image on the ultrasound display
		1.8	Adjust the patient and/or TRUS to acquire good image (avoid pubic arch interference) (optional)
		1.9	Once set, calibrate with the computer
2	Measure volume	2.1	Call out TRUS positioning planes in 0.5cm intervals
		2.2	Contour the prostate: mark each prostate slice on TRUS to measure overall prostate volume
		2.3	Call out the prostate volume read on ultrasound display to Medical physicist
		2.4	Print out each TRUS plane on TRUS printer
3	Needle implantation	3.1	Decide on the treatment plan for needle and seed insertion with Medical physicists
		3.2	Confirm the treatment plan for needle and seed insertion once decided
		3.3	Insert peripheral needles as per treatment plan
		3.4	Call out needle positioning to Medical physicists
4	Volume post needle	4.1	Mark the prostate slice on TRUS to measure overall prostate volume after the needle

Operator: Radiation oncologist			
Main task: Treatment delivery			
Plan: Do all in one sequence; do optional if necessary			
Operation	**Operation Description**	**Sub-operation number**	**Detailed sub-operation description**
			implantation
		4.2	Read out the volume post needle implantation
5	Virtual prostate and urethra positioning	5.1	Perform urethral catheterisation on the patient (use gel, saline, syringe, gloves, catheter, etc.)
		5.2	Highlight the urethra on the TRUS display to Medical physicist
		5.3	Confirm virtual prostate and urethra positioning with Medical physicist
6	Peripheral seeding and planning	6.1	Adjust the seed positioning based on the number of needles inserted with Medical physicist for optimal isodose by using the virtual 3D model (equal radiation intensity to the area (optional))
		6.2	Insert seeds via peripheral needles positioned as per treatment plan
		6.3	Call out the position of each needle and the number of seeds implanted to Medical physicists and to Nurse
		6.4	Review virtual seeds implanted and 3D virtual prostate model on computer display together with Medical physicist to determine best seed positioning (optional)
7	Central seeding and planning	7.1	Insert seeds via central needles positioned as per treatment plan
		7.2	Call out the position of each needle and the number of seeds implanted to Medical physicists and to Nurse
		7.3	Insert additional needle with seeds if necessary (optional)
		7.4	Call out the position of the additional needle and the number of seeds implanted to Medical physicists and to Nurse (optional)
		7.5	Review 3D model of virtual prostate on computer display together with Medical physicist
8	Volume post implant	8.1	Mark the prostate slice on TRUS to measure overall prostate volume after the needle implantation and to check for oedema
		8.2	Read out volume post seed implant to Medical physicists
		8.3	Remove the TRUS probe
		8.4	Dispose the needle template

Operator: Radiation oncologist			
Main task: Treatment delivery			
Plan: Do all in one sequence; do optional if necessary			
Operation	**Operation Description**	**Sub-operation number**	**Detailed sub-operation description**
		8.5	Remove the stepper
		8.6	Disassemble the bed
		8.7	Remove surgical gown, gloves, mask, cap, shoe covers & wash and disinfect hands
		8.8	Sign the seed evaluation form

THE DEVELOPMENT OF A TOOL TO EVALUATE SERIOUS INCIDENT INVESTIGATION REPORTS IN HEALTHCARE

C. McCaughan[1], S. Cromie[2], N. McDonald[3], and S. Hughes[4]

[1] *Lead Incident Management Investigation and Support/Researcher , National Incident Management and Learning Team, HSE Offices, Clonminch Business Park, Tullamore, Co. Offaly.*

[3] *Assistant Professor, Aerospace Research Psychology Group, School of Psychology, Trinity College Dublin.*

[4] *Associate Professor, Centre for Innovative Human Systems, School of Psychology, Trinity College Dublin.*

Abstract

An evaluation tool was developed for a study to determine the quality of reports of investigations of serious incidents carried out in a healthcare setting. The tool was tested for reliability prior to use to evaluate investigation reports that were completed in 2013. During the evaluation of the 2013 reports, opportunities to enhance the questionnaire to better achieve the objectives of the study were identified culminating in the development of an enhanced questionnaire tool for use for the evaluation of 2014 investigation reports. The next steps are for the evaluation tool to be used in May 2015 for the evaluation of serious incident investigation reports completed in 2014. Information from the experience of developing this evaluation tool has been and will continue to be communicated to the HSE for consideration in work which is ongoing to update the *HSE Guidelines for the Systems Analysis Investigation of Incidents and Complaints"* (2012).

Introduction

The reported rate of incidents occurring in health systems throughout the developed world is very high. Vincent *et al*., (2008) note that over 10 per cent of patients experience an adverse event while in hospital in the UK. This figure is comparable to other studies around the world. Landrigan *et al*., (2010) in the USA found adverse events to be as high as 25 per cent. McGlynn *et al*., (2003) in a study in the USA showed that only 55 per cent of patients received care deemed appropriate. Research by Kohn *et al*., (2000) in the USA suggests that of the over 33.6 million admissions into US hospitals in 1997 that 44,000 and possibly as many as 98,000 Americans died as a result of medical errors. The authors of the report conclude that more people die in the USA in a given year as a result of medical errors than from motor vehicle accidents or breast cancer.

Incidents are created by people operating within complex systems. These incidents can be analysed, and the lessons learned applied to prevent or reduce the chance of similar events recurring. The evidence suggests that where lessons are not learned similar incidents can and do recur (Toft and Reynolds, 2005).

It is a cornerstone of safety in any organisation that good quality internal investigations of incidents are undertaken in order to be able to identify and address the factors that contribute to harm.

Within a large and complex organisation such as the Health Service Executive (HSE), the implementation of the recommendations of individual incident investigation reports at the sites where the incidents occur is important to improve safety at that specific site. Aggregate analysis of all investigations of incidents that occur throughout the entire organisation over a given period is also important to gain an understanding of what causes most harm to most people most often and conversely to learn what national safety interventions are likely to have the greatest impact on safety for the greatest number of people the greatest amount of the time.

The *"HSE Guidelines for the Systems Analysis Investigation of Incidents and Complaints"* (2012) were developed to help ensure that investigations were as valid, reliable and generalisable as is possible. They are based on the "*Systems Analysis of Clinical Incidents: The London Protocol"* (Taylor-Adams and Vincent, 2004). While the London Protocol deals only with clinical incidents, the HSE Guidelines have been expanded to include non-clinical incidents. They were developed by the National Incident Management Team (NIMT)[1] following extensive consultation and engagement with service users, including service users that were harmed by healthcare; HSE staff and managers; regulators and other stakeholders; national and international general safety experts and patient safety experts; and based on extensive experience of managing and investigating incidents.

Purpose

The purpose of this paper is to report on the development of a tool to effectively collect data in relation to the quality of serious incident investigation reports including the following:

- Timeliness of investigations
- Quality of data collected for investigations
- Quality of chronologies reflected in investigation reports

- Quality of analysis reflected in investigation reports to identify key causal[2], contributory[3] and incidental factors[4]

- Quality of recommendations in relation to the use of a hierarchy of preferred control measures to develop recommendations; their linkage to contributory factors; and how SMART (i.e. Specific, Measurable, Achievable, Realistic and Time-bound) recommendations are

- The *"validity"*[5] and *"reliability"*[6] and the *"generalisability"* of investigation reports. Generalisability in this context is defined as the extent to which the findings of the enquiry are more generally applicable outside the specifics of the situation studied (Robson, 2002).

Method

Literature search and review

A review of the literature was conducted to identify tools to evaluate reports of investigations of serious incidents. No such tools were identified.

A tool entitled *"Investigation credibility and thoroughness criteria"* by the now terminated National Patient Safety Agency (NPSA) in the UK was identified through contact with users in the National Health Service (NHS) in the UK. The NPSA *"Investigation credibility and thoroughness criteria"* dealt with aspects of incident investigation and related definitions that had been modified for HSE Incident Investigation guidelines based on considerations, consultation and engagement as described in papers by McCaughan *et al* (2013a and b).

The HSE NIMT had developed an audit tool to monitor compliance with the "*HSE Guidelines for the Systems Analysis Investigation of Incidents and Complaints"* (2012) for use by investigators and by commissioners of investigations to ensure that investigations were conducted in line with the guidelines. The audit tool within these guidelines focused on collecting evidence that investigation reports complied with these guidelines. The format of the Audit tool required auditors to answer *"Yes/No/Partial"* to questions about whether investigation reports reflected compliance with the investigation guidelines and included a space for *"Comments"* in relation to each question. Users of the audit tool reported that the audit tool could reflect that an investigation report was compliant with the investigation guidelines but that there could be issues with the quality of the investigation report which the audit tool was not sensitive enough to identify.

The purpose of the current study required an evaluation tool to be developed which did not only collect data as to whether or not investigation reports complied with the guidelines; but it also needed to collect data in relation to the quality of the element of the investigation reflected in the various sections of the investigation report.

For example, the audit tool within the "*HSE Guidelines for the Systems Analysis Investigation of Incidents and Complaints"* (2012) asked the following question in relation to key causal factors:

> *"Are key causal factors identified and/or does the report state that none were identified following the analysis of the chronology?"*

The original questionnaire tool for this study asked the following question in relation to key causal factors:

"Where the equivalent ofkey causal factors have been identified, the investigation report shows appropriate analysis of the key causal factors or the equivalent to identify the underlying contributory factors orequivalent"	Strongly disagree	1	2	3	4	5	Strongly agree
Give details if necessary or if you give a score of 3 or less							
Additional comments							

Development of questionnaire

The questionnaire to evaluate reports of investigations of serious incidents was developed based on the NPSA *"Investigation credibility and thoroughness criteria"* and the audit tool in the "*HSE Guidelines for the Systems Analysis Investigation of Incidents and Complaints"* (2012) and based on consideration of the purpose of the study.

Questionnaire inter-rater reliability testing

Inter-rater reliability was tested qualitatively. Three randomly selected investigation reports were reviewed by three reviewers who then completed the questionnaire. Following this review the three reviewers met to discuss the completed questionnaires and any significant variations in responses between the three reviewers. The following two issues were considered at this meeting:

- For questions where Likert type scales were used, and where the responses by the reviewers varied by a magnitude of two or more, the reviewers would consider the questions with a view to enhancing the clarity and reducing the ambiguity of the question
- To check whether it was possible to get the answers for the questions from the completed investigation reports

There were no questions where the responses by the reviewers varied by a magnitude of two or more. It was identified that it might not be possible to answer some questions from the information available in the completed investigation reports such the question about whether contributory factors that could occur elsewhere were communicated to the relevant senior managers. However, it was agreed that these questions should be retained on the questionnaire in case some investigation reports might include the information.

It was also anticipated that the structure and content of the investigation reports might vary significantly and so it was considered that further review and enhancement of the questionnaire might be needed as more investigation reports were evaluated.

Finally, it was identified that, in the absence of published tools to evaluate the quality of investigation reports, and the related dearth of published criteria identifying how the various elements of an investigation report might be categorised - the *"additional comments"* section for each question would be very important to shape an enhanced questionnaire.

Use of the questionnaire to evaluate serious incident investigation reports

A single reviewer used the evaluation tool to review 61 serious incident investigation reports which were completed in 2013. It took an average time of 74 minutes to review each investigation report and to complete the evaluation tool.

At the start of the study, the original questionnaire for evaluating completed investigation reports was used and the data from the questionnaire was inputted onto an SPSS data base. The original questionnaire and the SPSS data base were enhanced with additional questions contemporaneously as investigation reports were being reviewed and when the review revealed that these questions needed to be included to achieve the objectives of the study. When all 61 investigation reports had been reviewed and evaluated, the reviewer re-evaluated the earlier investigation reports in line with the enhanced questionnaire and entered outstanding data onto the enhanced SPSS data base to ensure a complete set of data in line with the enhanced questionnaire. Please see details of the enhanced questions in the results section below. .

Results

Summary of changes made to the questionnaire following review of investigation reports

As anticipated, the structure and content of the investigation reports did vary significantly, posing challenges for analysing the reports. Learning from this phase of the study identified that it was not just important to refine the scales used to evaluate the reports, but that it was also necessary to develop the tool to better capture the details of the differences in actual reportsThe significant variation in the structure and content of investigation reports was also found to contribute to extended timelines in completing the reviews of investigation reports. Where investigation reports complied with HSE investigation guidelines including following the template investigation report that is available within these guidelines, it was possible to complete the review in a more timely manner. It was also identified that the HSE investigation report template could be enhanced to improve the *"analysability"* of investigation reports including the timeliness for reviewing investigation reports and completing the investigation evaluation tool that is the subject of this paper.

For each question, i.e. question about the quality of the chronology in the investigation report; question about whether appropriate data collection and review methods were used etc., it was identified that a "Yes/No" question about whether these elements were or were not reflected in the investigation report should be included.

Analysis of the qualitative data collected in the *"Additional Comments"* sections helped to define specific criteria about each element of the report. For example, the question on the original questionnaire related to the quality of data collection reflected in the investigation report was as follows:

"The investigation report shows appropriate data collection methods including collecting data from literature, records, interviewsetc"	Strongly disagree	1	2	3	4	5	Strongly agree

From the analysis of the data collected in the *"Additional Comments"* section, the following additional questions were added in a format that requested a yes/no response:

- Does the investigation report include a summary of the data collected and reviewed?
- Does the report reflect that the relevant records/healthcare records were collected and reviewed?
- Does the report reflect that the relevant local, national and international policies, procedures and guidelines were collected and reviewed?
- Does the report reflect a review of the relevant literature?
- Does the report reflect that individual interviews were conducted with individuals who observed the incident and/or had information pertinent to achieving the objectives of the systems analysis investigation?

This is now followed by a question to rate the quality of the data collected as "Excellent", "Very Good", "Good", "Fair", or "Poor":

Questions where it was identified that the information was not collectable from the investigation reports included the following:

- Date when decision to conduct investigation occurred
- Date when investigation report was released to those affected/harmed

Additional questions that were identified for inclusion relate to questions collecting information about the key causal factors, the contributory factors, and whether reports were anonymised.

Next steps:

Information from the experience of developing this evaluation tool has been and will be continued to be communicated to the HSE for consideration in work which is ongoing to update the *HSE Guidelines for the Systems Analysis Investigation of Incidents and Complaints"* (2012).

The evaluation tool will be used in May 2015 for the evaluation of serious incident investigation reports completed in 2014. Three reviewers will use the questionnaire to evaluate 10% of the investigation reports and the tool will be finalised for use then. A single reviewer will complete the questionnaire for all investigation reports. A second reviewer will complete the questionnaire for 25% of the investigation reports to check the reliability of the evaluation of the serious incident investigation reports.

Acknowledgements

The authors would like to acknowledge staff from the former HSE Regional Offices for Quality and Patient Safety; and the Offices of the Divisional Leads for Quality and Patient Safety who helped with collecting data for this study. The authors would also like to acknowledge members of the HSE NIMT/NIMLT who helped with collecting and compiling the data and with the reliability testing of the evaluation tool.

References

Health Service Executive. 2012, *Guidelines for the Systems Analysis Investigation of Incidents and Complaints.*

Kohn, L. Corrigan, J. and Donaldson, M. 2000, *To Err is Human: Building a Safer Health System*. National Academy Press, Washington DC.

Landrigan, C.P. Parry, G.J. Bones, C.B. Hackbarth, A.D. Goldmann, D.A. Sharek, P.J.. 2010, *Temporal Trends in Rates of Patient Harm Resulting from Medical Care*. New England Journal of Medicine; 363:2124-34 Nov 25, 2010

McCaughan, C. Cromie, S. McDonald, N. 2013a, *Factors in determining the level of causal analysis in healthcare incident investigations;* Proceedings of the Irish Ergonomics Society Conference & Symposium on Human Factors in Healthcare and Patient Safety. ISSN 1649-2102

McCaughan, C. Coyne, D. Hughes, S. Cromie, S. 2013b, *Enhancing system analysis in healthcare incident investigation: A work in progress;* Proceedings of the Irish Ergonomics Society Conference & Symposium on Human Factors in Healthcare and Patient Safety. ISSN 1649-2102

McGlynn, E. A. Asch S.M. Adams, J. Keesey, J. Hicks, J. DeCristofaro, A. Kwee, E.A. 2003, *The quality of healthcare delivered to adults in the United States*. New England Journal of Medicine: 348, 2, 635-45

Robson, C., 2002, *Real world research ; A resource for social scientists and practitioner - researchers.* Second Edition; Blackwell Publishing

Taylor-Adams, S. Vincent, C. 2004, *Systems Analysis of Clinical Incidents: The London Protocol.* Clinical Safety and Research Unit, Imperial College London.

Vincent, C. Aylin, P. Franklin, B.D. Holmes, A. Iskander, S. Jacklin, A. and Moorthy, K. 2008, *Is Health Care Getting Safer?* British Medical Journal, Vol. 337, No. 7680 (Nov. 22, 2008), pp. 1205-1207

Toft, B and Reynolds, S. 2005, *Learning from Disasters: a management approach* (Extended 3rd Ed).

PEWS: IT'S MUCH MORE THAN A SCORE - SUPPORTING CLINICAL JUDGEMENT AND SITUATION AWARENESS

Dr John Fitzsimons[1], Ms Rachel MacDonell[2], Dr Veronica Lambert[3], Prof Anne Matthews[3]

[1]*Clinical Director Quality Improvement Health Service Executive & Chair PEWS Guideline Development Group*
[2]*National PEWS Coordinator Health Service Executive*
[3]*School of Nursing and Human Sciences, Dublin City University*

Abstract

Paediatric Early Warning Systems (PEWS) include bedside tools which help alert staff to clinically deteriorating children. A specific priority of the National Clinical Programme for Paediatrics and Neonatology is the implementation of a National Paediatric Early Warning System. This paper presents findings from a systematic literature review and baseline pilot research both of which underpinned the development of new national age-specific paediatric observation charts that incorporate the Paediatric Early Warning System score, as well as a comprehensive educational package, to support the development of the National Clinical Guideline on PEWS for the Irish Health System. Key findings were recognising that PEWS is much more than a score and should never replace clinical judgement; instead through creation of a common language it should stimulate a heightened sense of situational awareness and open communication about children at risk of clinical deterioration.

Background

Many paediatric deaths are identified as either avoidable or potentially avoidable, with evident deterioration of symptoms (physiological and behavioural) often present in the 24 hours preceding a cardiopulmonary arrest. This provides a solid foundation for an increased attention to prevention; early detection through implementation of early warning scores and appropriate timely responses to the clinically deteriorating child. Paediatric Early Warning Systems (PEWS) include bedside tools which help alert staff to clinically deteriorating children by periodic observation of physiological parameters and predetermined criteria for escalating urgent assistance. A specific priority of the Irish Health Service Executive's National Clinical Programme for Paediatrics and Neonatology is the implementation of a National Paediatric

Early Warning System as recommended by HIQA's 2013 'investigation into the safety, quality and standards of services provided by the Health Service Executive to patients, including pregnant women, at risk of clinical deterioration, including those provided in University Hospital Galway, and as reflected in the care and treatment provided to Savita Halappanavar'. In 2014, chaired by Dr. John Fitzsimons, Consultant Paediatrician and Clinical Director for Quality Improvement HSE, the PEWS steering group developed new national age-specific paediatric observation charts that incorporate the Paediatric Early Warning System score, as well as a comprehensive supporting educational package and other resources. These developments were underpinned by the completion of a systematic literature review and baseline pilot research to support the development of the National Clinical Guideline on PEWS for the Irish Health System.

Aim

The aim of this paper is to twofold; firstly to, present the overall findings from the systematic literature review which assessed the evidence on the use, validation and education of early warning systems used in acute paediatric healthcare settings for the detection and/or timely identification of deterioration in children aged 0-16 years; and secondly, to offer insights from the evaluation of the pilot implementation of PEWS in four paediatric inpatient healthcare settings. A core concept to emerge from both the systematic review and pilot implementation was the importance of recognising PEWS as more than a score; rather it is one piece of a complex intervention supporting clinical judgement and situation awareness.

Systematic review methods

The methodology for the systematic review followed the Centre for Reviews and Dissemination guidance for undertaking systematic reviews in healthcare and the National Clinical Effectiveness Committee Guideline Development Manual with regard to considering evidence for the review; search methods; data collection and analysis including data extraction; quality assessment and data synthesis. A variety of electronic databases (i.e. Pubmed, Medline, Cinahl, Embase and Cochrane) and other resources (i.e. guideline clearing houses, grey literature databases, trial registers, professional organisations and association websites, and Google and Bing internet searches) were searched to retrieve published and unpublished evidence nationally and internationally; including clinical guidelines, primary research studies, secondary reviews, economic evaluations/analysis and grey literature. All searches, screening and data extraction was conducted by at least two reviewers with any discrepancies resolved with a third reviewer. Owing to the diversity of studies examining different aspects of PEWS, in reporting the findings we divided data extraction and narrative synthesis into detection systems, response systems, implementation processes, educational interventions and cultural, sociotechnical and organisational issues.

Systematic review findings

The systematic literature review (of 11 clinical guideline documents, 70 research papers and various sources of grey literature), identified that Paediatric Early Warning Systems are extensively used internationally in paediatric inpatient hospital settings; however there is no consensus and limited evidence about which PEW system is most useful or optimal for paediatric contexts. No robust national evidence-based clinical guideline to support the implementation of PEWS was identified. Definitive conclusions could not be made on the effectiveness of PEW systems for the detection and/or timely identification of, and response to, deterioration in improving clinical outcomes for children aged 0-16years in inpatient hospital settings. Some evidence did highlight positive directional trends in improving clinical based outcomes (e.g. reduced cardio-pulmonary arrests, earlier intervention and transition to PICU) for children who are clinical deteriorating, in addition to, reporting potentially favourable outcomes for enhanced multi-disciplinary team work, communication and confidence in recognising, reporting and making decisions about child clinical deterioration.

PEW detection systems

Diversity in PEW detection system physiological (and other) parameters, differences in age dependent vital sign reference ranges, and limited consensus on clinical deterioration outcome measures made it difficult to compare and contrast the performance criteria of PEW detection systems (e.g. Monaghan, 2005; Duncan *et al.* 2006; Parshuram *et al.* 2011). Although rare for any system to have both a high specificity and sensitivity, some scoring systems did show some promising sensitivity and specificity, however alongside considering validity of the scoring system many contexts chose simplicity and clinical utility as a priority in electing which PEW detection system score to implement (e.g. Monaghan, 2005; Duncan *et al.* 2006).

PEW response systems

Diversity existed in how institutions operationalised and evaluated the performance of PEW response systems such as rapid response/medical emergency teams (RRT/MET) with limited standardisation in adoption of a one or two tiered response system; team composition; activation/calling criteria and clinical and process outcomes measured; thereby making any comparative conclusions difficult (e.g. Bonafide *et al.* 2014; Panesar *et al.* 2014).

PEWS implementation processes

Diverse implementation approaches were identified; ranging from adopting social marketing principles to quality/performance improvement initiatives to chart reviews and pre-post implementation surveys, thereby making comparative evaluations difficult with no conclusions being drawn on what is the 'optimal' implementation strategy to use to influence changes in clinical/process outcomes (e.g. Lobos *et al.* 2010; Hayes *et al.* 2012).

PEWS educational interventions

Limited evidence was identified that specifically focused on educational aspect of PEWS; two structured interventions/packages were sourced – COMPASS and RESPOND (McKay *et al.*

2013; Tume *et al.* 2013). These packages favoured self-directed e-learning mechanisms and peer training models such as train the trainer, alongside short real-life problem-solving scenario based face-to-face sessions.

PEWS cultural, socio-technical and organisational issues

While hard evidence was limited a topical argument about the failure of PEWS transpired, perhaps as a realisation of the broader and complex nature of PEWS, that of health care cultural contexts; thereby recognising the many barriers and facilitators to PEWS implementation (e.g. Azzopardi *et al.* 2011). The integration of situational awareness interventions into healthcare 'patient safety/risk' forum to supplement early warning scores were highlighted as important for acknowledging the tacit knowledge of experienced clinicians (e.g. 'watcher' defined as gut feeling about a patient that is at risk of deterioration; briefings and huddles) (Brady and Goldenhar, 2013; Brady *et al.* 2013).

Pilot implementation of PEWS

Running concurrently with the systematic literature review, the PEWS guideline development group developed new national age-specific paediatric observation charts that incorporated the PEWS score, as well as a comprehensive supporting educational package. The first iterations of these newly developed age-specific paediatric observation charts and associated educational package were implemented in four pilot paediatric inpatient healthcare settings; two wards in two national children's hospitals and two paediatric units in two regional sites. One focus group interview (lasting approximately 1-1.30 hours) was conducted in each pilot site with multi-disciplinary participants ranging in numbers from 8-13 per each focus group. Each focus group interview was transcribed verbatim and data were analysed thematically. Findings are presented below.

Usability of PEWS

The introduction of PEWS was welcomed and viewed positively. PEWS was seen as easy to use and guided in what to input. It was liked for its structure and guidance around reporting. It was clear on the correct chart to use for the appropriate child. While it was relayed that it takes a few more seconds to calculate the overall PEWS score there was a general feeling that PEWS did not contribute to an increased workload (in some cases it resulted in less vital sign monitoring).

> *"I think everybody is liking the tool and liking the structure around reporting and who to go to and what to do when something isn't just quite right" (FG3)*

Benefits of PEWS in changing practice

A number of benefits related to the implementation of PEWS in changing practice were relayed, including; clearer documentation; patients seen earlier; standardisation of care delivery and creation of a common language; concrete evidence to support clinical judgement; forced to pay attention to a score and have to take action; helping less experienced staff think and respond; assisting managers in delegation and prioritisation of patient allocation/care; removal of routinized task of doing observations; enhanced focus on and prioritisation of patients needing close monitoring; enhanced structure and consistency in way staff communicate; opening collaborative discussions about child cases; improved inter-disciplinary communication; heightened awareness of the sickest children; getting a rapid response and an enhanced sense of urgency and improved safety.

> "*The first thing I do when I am on call to go and look at the boards to see if there are any kids on high PEWS scores so it highlights any kids who would be sicker and I start my shift and I might expect a call about them" (FG4)*

Factors to consider in implementing PEWS nationally

Focus group participants spoke about many factors to consider for the national roll-out of PEWS such as; adequate notice that/when PEWS implementation is happening; adequate 'roll-out time' for staff education; time for staff to read the COMPASS manual; avoiding too long a lead-in time so as not to lose momentum; implementing PEWS in a phased way throughout hospital/unit with supervision and support required at ward level and continuation of the auditing process.

> *"The challenge….and that is why we would recommend ….. that it be brought in in small slices in hospitals that are big like the paeds hospitals, it needs so much supervision and so much support for the girls on the wards" (FG3)*

PEWS education and training

Important elements to consider in relation to PEWS education and training suggested by participants were; practical scenario based sessions; multi-disciplinary teaching; more content emphasising clinical judgement and temporary escalation amendments; informal training opportunities for reflecting on practice such as auditing and grand rounds and considering opportunities for re-training.

> *"But I think definitely it needs to be multi-d and I think that everybody in the room needs to be hearing the same information so they are all taking it away…….I think it is important though because this tool is not just a nursing observation tool, it is a multi-d tool so it needs input from both" (FG1)*

Completion and interpretation of the PEWS scoring charts

A lot of time was spent with participants during the focus group to discuss the actual observation chart which incorporates PEWS. Key issues highlighted were; nurse and family concern; distinguish between expected and actual parent concern; how to assess family concern; cut off for low SpO2 %, recording FiO_2 accurately for different devices, interpreting effort of breathing, heart rate and blood pressure parameters and pictorial graph for temperature.

> *"I don't know there is also a difference between normal parental anxiety and...there is a difference between the anxious parent who …. again, I thought he was supposed to be on his*

antibiotics, what is the story here....... coming to you in all seriousness there is something seriously wrong with my child, I don't care what you say. And it is the same as the nurse feeling, not well, things aren't right. So that is concern" (FG2)

Amendments to PEWS parameters and temporary escalation

Amendments to PEWS parameters and temporary escalation amendments caused some discussion at focus groups around; which amendment to use – parameters or temporary escalation; more education on temporary escalation; how to document temporary escalation; timeframe for suspending escalation; use of clinical judgements with amended temporary escalation; and how to score PEWS when amendments are made.

"I think they have fears around it because when you talk them through it, the registrars, we say them, look it is a temporary escalation but you need to put in a timeframe and ...even in their whole language there is nearly a fear around it, do they have the ability to do that, do they have the confidence" (FG2)

Recognising the importance of clinical judgement

Participants across all four pilot sites emphasised that it is extremely important to highlight the fact that the PEWS numerical score is only one part of a holistic assessment of the child. PEWS should be combined with all other skills staff have, should not prevent anyone from taking action and should never be used to undermine someone's expertise and clinical judgement. Themes discussed included; PEWS sensitivity in capturing all 'sick' children; fear of numerical focus taking precedence over nurse concern; PEWS not for sole purpose of getting doctors to attend to the child; importance of escalating and responding to nurse concern; being situationally aware and opening discussions about a sick child; false reassurance of a low PEWS score and keeping observation skills to fore in educating future generations.

"It's about the awareness of the situation rather than just the score. So you are saying as an experienced nurse, we are not happy, or the nurse in question wasn't happy. That should have been listened to, I mean there is no doubt in the world about that and regardless of a PEW scoring X, Y or Z, you say you are concerned that the baby or child needs to be reviewed. If you rang any consultant in the world they would say, nurse worried, baby or child must be reviewed" (FG4)

Conclusion

PEWS should never replace clinical judgement; instead through creation of a common language it should stimulate a heightened sense of situational awareness and open communications about children at risk of clinical deterioration. The involvement of clinical stakeholders assisted guideline developers in re-designing the new national age specific paediatric observation charts, as well as the comprehensive supportive educational package. It is anticipated that national roll-out of PEWS to all paediatric inpatient healthcare settings will commence in September 2015. This is one part of a larger safety network which will develop and grow over time.

Acknowledgements

Both the systematic literature review and conduct of the baseline pilot research to support the development of a National Clinical Guideline on PEWS for the Irish health system were funded by the Clinical Effectiveness Unit, Department of Health and overseen by the National Clinical Effectiveness Committee and Health Service Executive PEWS Guideline Development Group.

References

Azzopardi, P., Kinney, S., Moulden, A., Tibballs, J. 2011, Attitudes and barriers to a Medical Emergency Team system at a tertiary paediatric hospital, *Resuscitation,* **82**, 167-174.

Bonafide, C.P., Localio, R., Roberts, K.E., Nadkarni, V.M., Weirich, C.M., Keren, R. 2014, Impact of Rapid Response System Implementation on Critical Deterioration Events in Children, *JAMA Pediatrics*, **168**, 25-33.

Brady, P.W., Muething, S., Kotagal, U., Ashby, M., Gallagher, R., Hall, D., Goodfriend, M., White, C., Bracke, T.M., DeCastro, V., Geiser, M., Simon, J., Tucker, K.M., Olivea, J., Conway, P.H., Wheeler, D.S. 2013, Improving situation awareness to reduce unrecognised clinical deterioration and serious safety events, *Pediatrics,* **131**, e298-e308.

Duncan, H., Hutchison, J., Parshuram, C.S. 2006, The pediatric early warning system score: A severity of illness score to predict urgent medical need in hospitalized children, *Journal of Critical Care*, **21**, 271-279.

Hayes, L.W., Dobyns, E.L., DiGiovine, B., Brown, A., Jacobson, S., Randall, K.H., Wathen, B., Richard, H., Schwab, C., Duncan, K.D., Thrasher, J., Logsdon, T.R., Hall, M., Markovitz, B. 2012, A Multicenter Collaborative Approach to Reducing Pediatric Codes Outside the ICU, *Pediatrics,* **129**, e785-e791.

Lobos, A.T., Costello, J., Gilleland, J., Gaiteiro, R., Kotsakis, A., on behalf of the Ontario Pediatric Critical Care Response Team Collaborative. 2010, An implementation strategy for a multicentre pediatric rapid response system in Ontario, *The Joint Commission Journal on Quality and Patient Safety,* **36**, 271-280.

McKay, H., Mitchell, I., Sinn, K., Mugridge, H., Lafferty, T., VanLeuvan, C., Mamootil, S., Abdel-Latif, M.E. 2013, Effect of a multifaceted intervention on documentation of vital signs and staff communication regarding deteriorating paediatric patients, *Journal of Paediatrics and Child Health,* **49**, 48–56.

Monaghan, A. 2005, Detecting and managing deterioration in children, *Paediatric Nursing,* **17**, 32-35.

Panesar, R., Polikoff, L.A., Harris, D., Mills, B., Messina, C., Parker, M.M. 2014,

Characteristics and outcomes of pediatric rapid response teams before and after mandatory triggering by an elevated pediatric early warning system (PEWS) score, *Hospital Pediatrics*, **4**, 135-140.

Parshuram, C.S., Duncan, H.P., Joffe, A.R., Farrell, C.A., Lacroix, J.R., Middaugh, K.L.,

Hutchinson, J.S., Wensley, D., Blanchard, N., Beyene, J., Parkin, P.C. 2011, Multicentre validation of the bedside paediatric early warning system score: a severity of illness score to detect evolving critical illness in hospitalised children, *BioMed Central Critical Care,* **15**, R184.

Tume, L.N., Sefton, G., Arrowsmith, P. 2013, Teaching paediatric ward teams to recognise and manage the deteriorating child, *British Association of Critical Care Nurses,* **19**, 196-203.

STOP CALLING ME: IMPROVING COMPLIANCE WITH THE NATIONAL EARLY WARNING SCORE SYSTEM'

Paul O'Connor[1], Sinéad Lydon[2], Gozie Offiah[3], Louise Gleeson[3], Dara Byrne[4].

[1]Department of General Practice, National University of Ireland, Galway; [2]School of Psychology, Trinity College, Dublin, Ireland; [3]Beaumont Hospital, Dublin, Ireland; [4]Saolta Hospital Group, Galway, Ireland.

Introduction

Physiological track-and-trigger warning systems (PTTSs) are increasingly being used to aid the early identification of patients at risk of deterioration. PTTSs use periodic observation of selected basic vital signs (the 'tracking') with predetermined criteria (the 'trigger') for requesting the attendance of more experienced staff. PTTSs have generally been shown to positively impact patient outcomes such as mortality and the occurrence of serious adverse events[1]. However, research has also demonstrated that PTTSs are often poorly or improperly implemented, that staff may fail to adhere to the escalation protocols surrounding the PTTSs, and that the utility of PTTSs in resource deprived hospitals may be compromised[2,3].

In Ireland a PTTS called the National Early Warning Score System (NEWS) was introduced in 2012. Adapted from the Vitalpac™ Early Warning Score (ViEWS), this PTTS requires the observation of patients' respiratory rate, oxygen saturation and oxygen inspiration, heart rate, blood pressure, temperature, and level of consciousness[4]. Clinical guidelines dictate appropriate observation of patient vitals, escalation of patient care, and emergency response systems. Clinical communication is facilitated by a structured communication tool based upon ISBAR[5]. The current research study used a mixed-method approach to examine nurses' and doctors' attitudes towards NEWS and to identify factors that impact upon the intention of doctors and nurses to use the system.

Qualitative Phase

Methods: Qualitative.

During the initial qualitative phase, 30 nurses and doctors participated in a series of semi-structured interviews, developed from the Theory of Planned Behaviour (TPB).

The TPB is a psychological model intended to explain and predict human behaviour. It suggests that engagement in any behaviour is predicted by a persons' intention to engage in the behaviour. The model explains behavioural intention (*intention*; i.e., do you intend to continue to apply the NEWS as you have been doing, or do you think you may make changes in the future?) as an outcome of the social perception of the behaviour (*subjective norms*; i.e., do you think there are variations between how the NEWS is being used and interpreted by nurses and doctors?), an individual's own perception of the behaviour (*attitude*; i.e., what do you think the effects are of the NEWS on patient safety and quality of care?), and the individual's perception of whether they can perform the behaviour (*perceived behavioural control*; i.e., what factors hinder you from using the NEWS?).

Interview participants were recruited using a snowball technique. Interviews were conducted on-site at an Irish teaching hospital and lasted between 15 and 20 minutes. Appropriate ethical approval was obtained prior to recruiting the participants. This data was analysed using a deductive content analysis approach using the TPB constructs to theme the interviews.

Results: Qualitative.

Interviewees reported a positive impact of the NEWS on patient safety. Improvements in the quality of patient care that were reported included: more frequent monitoring of patient vitals; prioritisation of deteriorating patients; detection of ill patients more quickly and reliably; the increased likelihood of ill patients seeing a senior doctor, and early detection of sepsis.

The interviews also highlighted a number of barriers to the implementation of NEWS. These barriers included:

- Interns frequently cited the NEWS as a source of conflict between doctors and nurses and had a negative effect on teamwork;
- NEWS implementation is hindered by staffing levels;
- all hospital staff have not received training in NEWS;
- this system reduces the autonomy of nurses and does not allow for their use of clinical judgement;
- registrars and consultants rarely adjusted the parameters for patients whose baseline values fall outside of norms so that their condition can be effectively monitored; and

- participants suggested a better system relating to the charting of parameters be developed.

Quantitative phase

Method: Quantitative

A 27-item TPB questionnaire was developed based upon the data collected from the semi-structured interviews using the process outlined by Fishbein and Ajzen[6]. Responses to the items were in the form of a five-point Likert scale from 0 (disagree strongly) to 5 (agree strongly). The questionnaire was comprised of five subscales: subjective norms (6 items), attitudes (6 items), perceived behavioural control (3 items), behavioural intention (2 items), and perceived barriers (10 items). Information on participant's training in the use of the hospital's PTTS, their medical speciality, and their position at the hospital was also solicited. The questionnaire was distributed via email to doctors and nurses at two Irish teaching hospitals.

Results: Quantitative

A total of 215 nurses and doctors (24.1% response rate) responded. Table 1 shows the Cronbach's alpha scores for each subscale and subscale scores separated by respondents' position at the hospital.

Table 1. Summary data for the questionnaire subscales.

Sub-scales	**Alpha**	**Nurse**		**Intern**		**Senior non-consultant hospital doctor**		**Consultant**	
		Meanper item	**SD**	**Meanper item**	**SD**	**Meanper item**	**SD**	**Meanper item**	**SD**
Subjective norms	0.63	3.87	.80	3.21	.83	3.02	.71	3.81	.71
Attitudes	0.91	4.22	.82	3.31	.98	2.71	1.06	4.29	.67
Perceived behavioural control	0.82	4.57	.73	4.07	.93	3.32	.99	3.86	1.14
Intention	0.53	4.16	1.03	4.30	.86	2.95	1.20	3.63	1.16
Barrier	0.78	2.77	.79	3.34	.58	3.41	.70	3.11	.81

The Kruskal-Wallis test identified significant differences between the groups on each of the five subscales: *subjective norms*, $\chi 2$= 48.88, p<.001, η^2=0.25; *attitudes*, $\chi 2$= 70.49, p<.001, η^2=0.36;

perceived behavioural control, $\chi 2= 54.32$, $p<.001$, $\eta^2=0.28$; *behavioural intention*, $\chi 2= 40.19$, $p<.001$, $\eta^2=0.21$; and *barriers*, $\chi 2= 22.45$, $p<.001$, $\eta^2=0.12$. To identify where these significant differences lay among the four groups, follow-up tests were conducted.

For both the *subjective norms* subscale and the *attitudes* subscale, nurses and consultants did not differ significantly from one another but demonstrated significantly more positive attitudes towards the NEWS than interns and senior NCHDs. Nurses were observed to report significantly higher levels of *perceived behavioural control* than nurses, interns, and consultants while interns reported significantly higher perceived behavioural control than senior NCHDs. Senior NCHDs reported a significantly lower *intention* to use the PTTS than nurses and interns. Finally, nurses rated the *barriers* listed as significantly lesser than did interns or senior NCHDs. The other groups did not differ significantly.

Discussion

Overall, Irish healthcare professionals believe that NEWS has a positive impact on patient safety and quality of care. However, a number of barriers and issues with the system were identified:

- NEWS appears to have a negative impact on teamworking. In order to foster teamworking, it is suggested that staff are trained as a multi-disciplinary team, and that there is ongoing observation and feedback in order to encourage the appropriate application of the system.
- Staff sometimes adhered to NEWS protocol even when it conflicted with their own clinical judgement. This 'blind' adherence to protocol is likely related to the need for a 'just culture' in which staff are not blamed for 'honest errors'[7].
- A lack of participation by senior doctors was arguably the largest barrier to the implementation of NEWS. Consultants and registrars could be more proactive in adjusting the NEWS parameters on a patient-by-patient basis to prevent junior doctors being continuously called to see a patient whose condition has not worsened.
-

Conclusions

These findings reveal that there are a number of significant barriers to the implementation of NEWS. Implementation requires a measured, large scale and deliberate process, and continued regular observation and feedback from, and to, the users[8]. If the sociocultural aspects of the implementation of the PTTSs fail to be addressed, the risk of failing to recognise deteriorating patients will remain.

References

Alam N, Hobbelink EL, van Tienhoven AJ, van de Ven PM, Jansma EP, & Nanayakkara PWB. The impact of the use of the Early Warning Score (EWS) on patient outcomes: A systematic review. Resuscitation 2014;85:587-594.

Donohue LA, & Endacott R. Track, trigger and teamwork: Communication of deterioration in acute medical and surgical wards. Intensive Crit Care Nurs 2010;26:10-17.

Inpatient nursing care and early warning scores: A workflow mismatch. J Nurs Care Qual 2014;29:215-222.

National Clinical Effectiveness Committee (NCEC). National early warning score: National clinical guideline no. 1. Ireland; 2013. Available from: http://www.hse.ie/eng/about/Who/qualityandpatientsafety/Standards/nationalstandards/Nstandardsdocuments/nearlywarnibngscoreno1.pdf

Health Service Executive (HSE). Effective clinical communication: ISBAR (Identify Situation-Background- Assessment- Recommendations). Ireland; 2013. Available from: http://www.hse.ie/eng/about/Who/qualityandpatientsafety/Local_Quality_and_Patient_Safety_Offices/QPS_DNE/Tool-box_Talks/ISBAR.pdf

Fishbein M, Ajzen I. Predicting and changing behavior: The reasoned action approach. New York; 2010: Psychology Press.

Reason J. Human error: models and management. BMJ 2000;320:768-770.

Bosk CL, Dixon-Woods M, Goeschel CA, & Pronovost PJ. Reality check for checklists. Lancet 2009;374:444-445.

Systemic Human Factors for Ultra Safe Aviation

N. Mcdonald, M. Ward, R. Morrison, S. Corrigan, M. C. Leva, J. Cahill, A. Kay

Centre for Innovative Human System,
School of Psychology Trinity College Dublin
Ireland

Abstract

There can be many benefits associated with applying an appropriate strategy to guide an organisation into the future. A common theme amongst authors of strategic performance management material is often the degree of importance an organisation associates with its ability to measure its performance as informed by its organisational strategy. Indeed Kaplan and Norton (1992) who are credited with the development of a strategic performance management template (namely the balanced scorecard) compare a company's journey to the complex task of navigating and flying an aircraft.

1. Background

The Centre for Innovative Human Systems (CIHS) is part of the School of Psychology, Trinity College Dublin. It is an aviation Human Factors / Ergonomics (HF/E) research group in Europe with a strong focus on aviation operations as well as design of new technologies – it has a unique systemic approach to aviation safety. The CIHS (originally APRG - Aerospace Psychology Research Group) coordinated a series of high profile EU collaborative projects on aviation maintenance from 1995-2005. The R&D approach comprises two complementary strands of work: Design for Operations concerns the operational impact of new technologies; Managing Risk and Change concerns performance and change across the aviation system. These are converging in new applications that draw from the complementary research and development in both strands. The key actions performed by the Centre include the normal ways of doing tasks (as opposed to the official ways) and the difficulty of achieving change. This researchstrongly influenced the JAA regulations (Amendment 5 of JAR145/EASA Part145). From 2005-9 the APRG co-ordinated HILAS, one of the largest HF/E project in aviation (39 partners in 16 countries) that established a new HF/E framework for comprehensive HF/E design evaluation, operational performance management in flight operations and maintenance; and management of change (McDonald 2014). This also led to the development and delivery of HF/E training in maintenance through the STAMINA programme - an industry benchmark.
In aviation there are new challenges to be faced. New technology is driving integration – smart cockpits, 4-D trajectories, airport collaborative decision-making, system-wide-information-management (SWIM) – as integration creates new system constraints and dependencies, the role of people becomes more, not less, critical. New business models are driving operational norms to the boundaries of the regulations – but what are the real risks? Aviation is 'ultra-safe' – thus accidents are not just re-combinations of known contributory factors, but involve interactions of factors previously considered safe. How can we anticipate and control such systemic risks?

The approach proposed by the CIHS is to model the social and technical system, to analyse systemic risks, to design future technologies against operational criteria, to manage performance and to successfully change existing systems. The aim is to sustain value in a distributed system, supporting the authority and competence of crew and staff to act in an accountable way not only locally but also up to the highest level of aviation regulation.

2. Managing Risk and Change

If there is to be a verifiable performance improvement in aviation system safety then three conditions have to be satisfied:

- The management of risk and safety has to be integrated across the system, because of interdependencies between flight & ground operations, maintenance, and ATM.
- The precursors of variations in normal performance that could give rise to rare serious events need to be controlled.
- Reliable system change including collaboration between organisations that normally compete has to happen.

<u>Integrated risk:</u> The centre helped to develop a comprehensive set of safety performance indicators for SAS; these were normalised across flight operations and maintenance to ensure a common risk concept (e.g. significance of deferred defects). With Alitalia and Aeroporti di Roma the centre is now developing an integrated risk framework between an airline and an airport and linking it with a simulation study to show how this can be extended to ATM. This common dataset is being tested in a relatively simple case study – the prediction of bird-strikes – now moving on to more complex patterns involving characteristics of crew, aircraft, route, weather, etc.

<u>Innovative predictive system risk analysis</u>: This analysis provides a formal quantification of hazard likelihoods by detecting robust configurations of co-occurring independent predictor variables. The results confirm that risk patterns are detectable by Data Science. The pattern can then be broken down to remove the combination of multiple independent predictors and reduce the likelihood of the event criterion occurrence or exposure. Breaking the pattern is the new way to manage causal systems leading to very rare accidents or safety events.

<u>Managing Change</u>: Implementation of Airport Collaborative Decision Making (ACDM). 35 major European airports are mandated to implement ACDM within the next year. Arlanda airport is one of the airports with a wide range of independent companies whose active collaboration is necessary to make ACDM work successfully. CIHS has worked at Arlanda with Swedavia and their airport partners in analysing the human and social aspects of collaboration. This goes much further than other research initiatives including SESAR. A dedicated training and learning programme, based on this is now being delivered. Working with EUROCONTROL, we have targeted 20 major European hub airports as potential customers for implementation (Corrigan et al 2014).

In Pescara airport in collaboration with SAGA (airport company) CIHS have developed and implemented a novel performance management IT system comprising the Daily Journal (log of daily activities, based on comprehensive process mapping), Anomalies Report, and Hazard ID (Leva et al 2014).

Implementation has been very successful enabling the company to improve performance despite staff reductions. 200+ anomalies have been reported and solved. This has supported SMS compliance, culture change (open communication, willingness to report, contribute to solving problems), and is important in supporting the company's new business plan.

These elements comprise key components for proactive, systemic SMS. A workshop on current and future regulations will be hosted by EASA in June 2015.

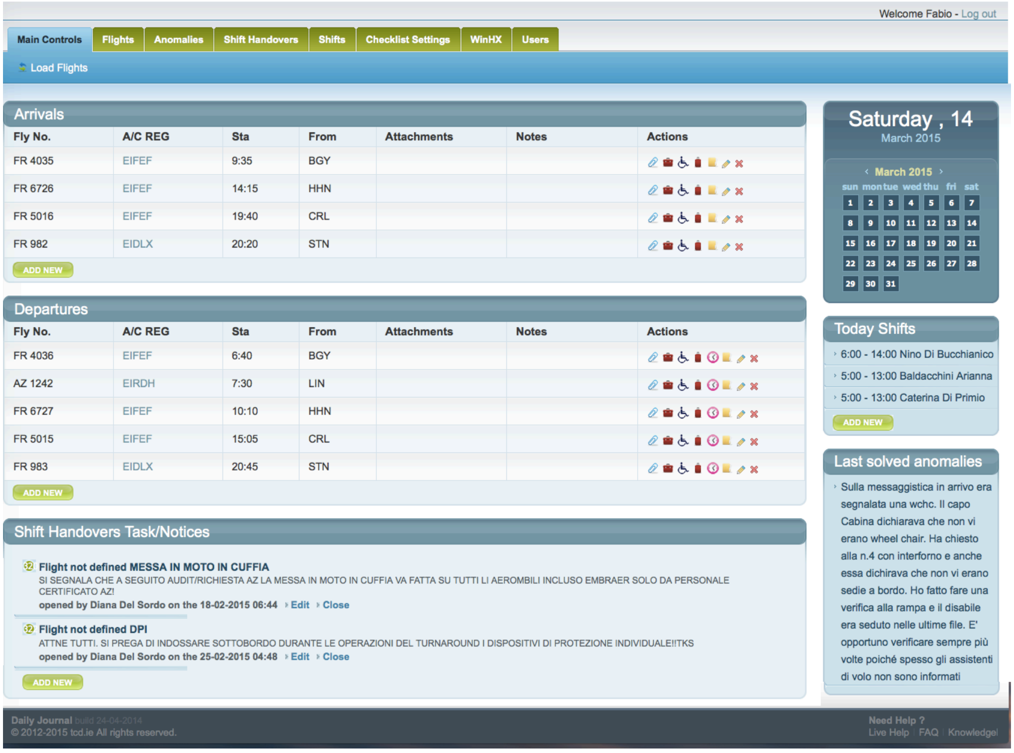


Figure 1. Screenshot of the Daily Journal as developed for Pescara airport

3. Design for Operations

Human factors research carried out in the centre has a key role in two strategic EU projects designing future aviation technologies. It leads the Human Factors integration of ACROSS (co-ordinated by Thales, with Airbus, Dasault, BAE Systems among 35 partners) developing advanced cockpit technologies to reduce peak workload and stress, supporting crew and managing pilot incapacitation. In A-PiMod (led by DLR, the German Aerospace Research Institute) CIHS leads the safety impact and validation across all stages of the project. This project is developing a new adaptive automation concept based on Multi-Modal Pilot Interaction, Operator Modeling, and Real-Time Risk Assessment.

Challenges concern demonstrating the operational impact of diverse technologies which are not yet fully integrated and may be hosted in multiple test-beds in different research centres. This requires co-ordinating diverse human factors research partners using multiple methods.

A HF/E Dossier provides a common framework to integrate the diverse HF/E methodologies, both at the local technology development and global operational level. The dossier spans scenarios, task modeling, HMI, operational dependencies, crew role in managing the operation, workload assessment and overall system risk assessment.

Integrated HF/E Teams deliver an ecological analysis of how technologies can change operations, using a mix of methods, combining formal and informal knowledge. A collaborative, participatory approach to design (e.g. collaborative prototyping) includes a stakeholder approach to evaluation through an External Expert Advisory Group and Community of Practice,

A powerful crew concept was developed based on crew maintaining a 'situation awareness bubble' (staying mentally ahead – or behind – of where they should be). At any point this crew are looking ahead (plan, anticipate), managing the current situation (monitoring, analysing, deciding, acting, checking) and reviewing what has happened up until now. This is constantly updated (See Figure 2).

In ACROSS, this crew concept enabled a powerful analysis, showing three distinct ways in which workload can be managed: reducing demand on the pilot through classic automation, monitoring crew physiological and behavioural activity to stimulate an appropriate reaction (up to taking control of the aircraft from the ground) and using smart automation to shift the balance of workload through better planning and anticipation. We have been able to demonstrate to the technology developers how to evaluate the revolutionary potential of their technology. As demonstrated by both projects, these new automation concepts bring the right information to the crew earlier in the process so that the crew has increased overall control of the situation and can deploy adaptive automation functions in a powerful manner. This in turn suggests how the next generation of applications should be constructed.

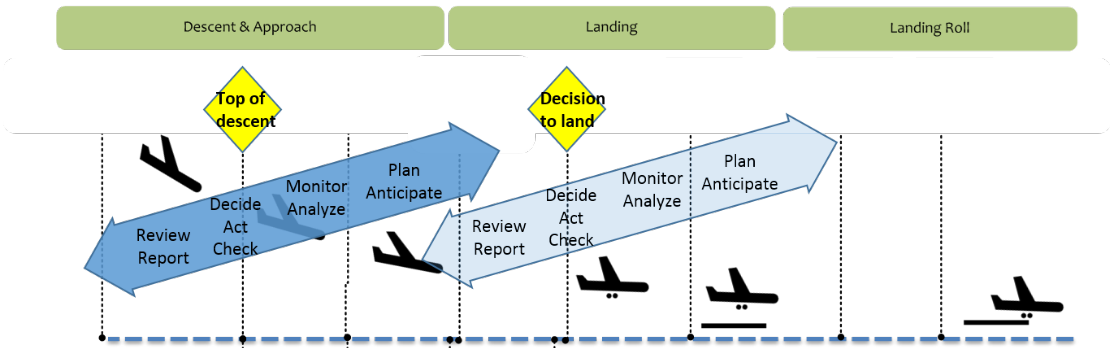


Figure 2. Crew actions for situational awareness in different flights phases (Apimod 2015).

Conclusion

There is a converging logic of system design and system change, making it possible for Human Factors research to play a key role in enabling the development of future aviation systems. New aircraft technologies and new IT systems create an explosion of information that can be integrated and transformed to create knowledge services to support flight crew, airport and maintenance staff, and ATC management through better system design and management of performance and change in an ultra-safe, accountable, distributed aviation system. Our current challenge is to intensify the transfer of R&D outcomes into successful innovation.

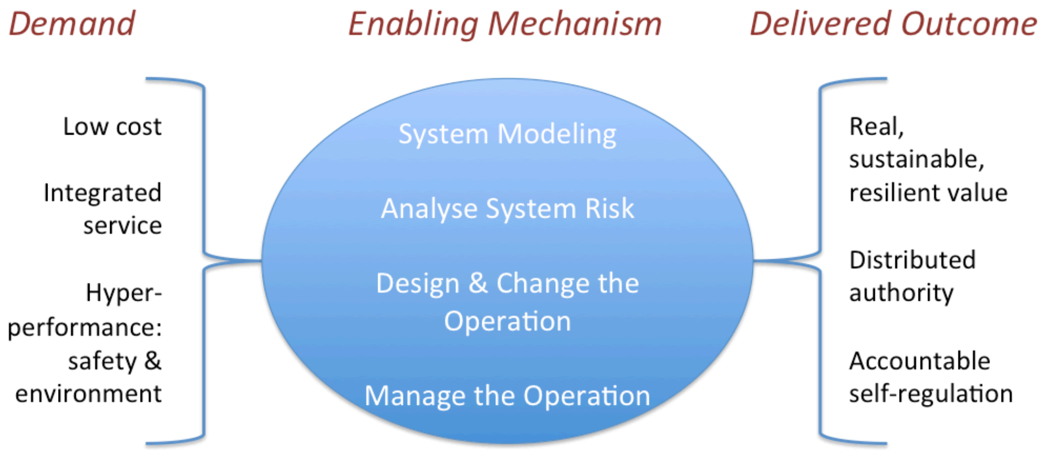


Figure 3. System and Environment; Demand, Capability and Response

Figure 3 integrates the ideas in this paper and sets out a concept of system and environment, with demands being addresses by socio-technical enabling mechanisms to produce appropriate outcomes. These outcomes are the principles through which the demands of a turbulent environment could be met. Thus the framework outlined in this paper sets out to provide the basis for the design, operation or change of systems which have real, sustainable and resilient value, distributed authority and accountable self-regulation.

References

APIMOD 2015 Official website fo the Project www.apimod.eu (last accessed 10th of June 2015)

McDonald, N (2014): *The evaluation of change*. Cognition Technology and Work 16, (4), 2014 Journal Article.

Corrigan, S., Mårtensson, L., Kay, A., Okwir, S., Ulfvengren, P. (2014). Implementing Collaborative Decision Making: Challenges & Recommendations. Journal of Cognition, Technology & Work (accepted for publication for Special Issue on Managing Change due for publication September 2014)

Leva, M.C., Del Sordo, D., Righi, P., Mattei, F (2014) Performance Management in a Small Regional Airport, from Day to Day Data Collection to an integrated SMS. Journal of Cognition, Technology & Work (accepted for publication for Special Issue on Managing Change due for publication September 2014

USING WEARABLE SENSORS TO MEASURE POSTURE

Bridget Masterson

Intel Ireland, Leixlip, Co. Kildare

Abstract

There has been an explosion of wearable devices in the marketplace in the last few years. Many of these have been directed at individual consumers in the lifestyle and health arenas. Low cost, user friendly, accurate, real-time data collection on posture and movement for professionals has not been possible until recently. Historically, the Intel Ergonomist used multiple tools (observational skills, goniometer, stopwatch, clicker, pen and paper) and needed to be in close proximity to the subject to observe postures. There was no methodology to record the continuous postures adopted during a task. The traditional methods are open to significant inaccuracies, particularly when employees are wearing fab bunny suits as body angles are hidden by the bunny suit. At Intel, two wearable systems have recently been implemented. System 1 is a wearable sensor system that can be attached to different parts of the body making real -time, accurate and repeatable measurements of how the body moves. System 1 sensor s use specifically designed visualization software to provide the ergonomist with live, animated feedback of the movement of the body including the head, neck, shoulder, forearm, elbow, wrist and hip. The data from System 1 can be easy transferred via Bluetooth to a tablet. The System 1 sensors will allow accurate data to be used during the ergonomic risk assessment process on the factory floor. The second system consists of two products, a back module and a lift module. The back module is a belt worn around the waist. The lift module is a posture-tracking wearable sensor worn as a small button near to the collar bone. Both track data such as good posture, steps, distance travelled, and calories burned. There is a vibration alert mode which is triggered when the user adopts a non-neutral posture. The lift module has a coaching function, which attempts to train people into the habit of sitting in a neutral position. The data from System 2 can be viewed on a PC or smartphone. System 2 can be used proactively by off ice-based employees who wish to improve their overall posture and health but who currently lack data or motivation to do so.

WHY USE A BAYESIAN ANALYSIS FOR GIGANTIC DATA SETS IN SAFETY RESEARCH?

T. Waldmann and A. McCormack

Department of Design and Manufacturing Technology
University of Limerick
Castletroy, Limerick
Ireland

Abstract

In her PhD thesis McCormack (2014) defended her view that large data sets should not be analysed using Statistical Hypothesis Inference Testing (SIT). She analysed data for all US Navy shipyard injuries and safety deficiencies and demonstrated that using common methods of statistical inference found in SPSS, for example, would lead to a successful capture of asterisks, but not to meaningful interpretations. "Significant" differences and relationships can be found in compilations of data points if there are sufficient numbers of them, when frequentist statistics are used.

The 250-year-old spat between frequentist and Bayesian statistics has been more than an esoteric debate on the nature of probability. This paper explains how approaches to statistics have influenced jury decisions in paternity cases and childcare as well as in pharmaceutical research and academic careers.

For example, since last year, the journal Basic and Applied Social Psychology has banned the use of p-values. In a 1990 Nature article, Blinkhorn and Johnson argued that if we stopped culling tabular asterisks and ignored effect sizes "below the ambient noise". Personality Testing would disappear! A p-value is often misunderstood to mean that the Null Hypothesis is probably true (or false), given the data, or, (very popular this!) that it is the probability that the results occurred due to chance. After discussing the dichotomous thinking of SIT and rejecting it in most cases, three options exist for the serious researcher: manipulate the experimental contingencies in such a way that each participant is his or her own control, measure effect sizes and power of tests, as is common in medicine and officially demanded by American Psychological Association journals, conduct meta-analyses, or use Bayesian statistics. The latter was the approach of choice for data that are voluminous, and which had been collected by various sources in various settings, with inevitably varying operational definitions. In applying this method, we continually update the probability of a hypothesis as evidence is acquired. Bayesian updating is particularly important in the dynamic analysis of a sequence of data, data that are found in the safety and injury data the candidate analysed.

THE ROLE OF HUMAN FACTORS FOR TOTAL SAFETY MANAGEMENT APPLICATION: THE TOSCA PROJECT

M. Chiara Leva[1], Nora Balfe[1], Tom Kontogiannis[2], Marko Gerbec[3],

[1]Centre for Innovation in Human Systems – Trinity College, Dublin, Ireland
[2]Department of Mechanical Engineering– Technical University of Crete Greece
[3]Joseph Stephan Institute Ljubljana Slovenia

Abstract

Over the recent past the accumulation of major mishaps, crises and accidents have made it clear that organisations must still improve their capabilities to address safety "not a stand-alone activity that is separate from the main activities and processes of the organization" but as an integrated part of critical project management. Further it is critical also to understand how weaknesses not only in the technical but also in the organisational interfaces can contribute to significant losses and major industrial accidents.

TOSCA (Total Operation Management for Safety Critical Activities) is concerned with the integration of industrial operations into a total performance management system so that concerns about safety, quality and productivity are addressed in an integrated way during lifecycle of a project or a product. The industrial domain of application regards process control industries (e.g., chemical industries, power generation, offshore oil & gas platforms, etc.) that may vary in size, regulatory and cultural aspects. Tosca has examined vulnerabilities of the technical, human and organisational systems that may have an impact in safety, quality and productivity. Safety critical activities can be seen as 'projects' or 'safety cases' that must be examined from the perspectives of many stakeholders (e.g., different departments, subcontractors, regulatory authorities, etc) and decision-making at different organizational levels (e.g., top managers, supervisors and operators). A participative approach is required that should collect knowledge from the sharp-end operators and integrate it with formal descriptions of how the system works and how responses should be coordinated across the whole organisation. Furthermore, TOSCA tackled also the issue of management of changes not just from the technical point of view but also considering human and organizational aspects. The project has been testing out the effectiveness of the methods and tolls developed so in a number of test beds. The project was able to provide:

- Identifications of the industry needs for a new integrated methodology, that will comprise safety, quality and operations management.
- Definition of the basic elements of the new integrated methodology, based on existing methods for safety, quality and operations management.

- Provision demo and specifications for the software tools, used to support the new methodology, based on industrial needs.
- A number of case studies where the outcome of Risk assessment methods informed by Human and Organizational factors were able to deliver improvements not only in terms of Health and Safety but also in terms of process optimization, quality and productivity.

The modular set of methods and tools are being tested on a variety of industrial applications both for Small Medium Enetrprises (SMEs) and Lager enterprises some examples are: an LPG storage and a distribution site, a company that produces agro-chemical products, an Electricity production and distribution company etc.

Most of the methods and tools were organized to respond to concrete industrial needs, the overview may seem complex but in the end it was about solving one problem at the time as for example:

1) The issue of building a common operational picture. This stem from the need to build on a set of existing and new methods of risk assessment in order to improve the identification of hazards and barriers for risk mitigation linked to everyday operations and tasks. This aspect for instance has been applied to an SME that produces malt to improve their capacity to identify and monitor actual risk on their day to day operations and their barriers. Another example is the one developed for the Agrochemical company where the data from the existing HAZOP was used to build an overview of the main accidental scenarios and identify the barriers or safety critical activities for which better procedures and monitoring was required.

Similarly moving to the area of supporting safety by design the project team used the mapping of main processes and procedures to feed into the possibility of redesigning a plant or a procedure using rapid prototyping to visualize the task at hand and support the identification of possible issues and the optimization of the entire process.

This same process has been used to feed into the revision of critical activities: such as the loading and loading tasks, that lead to the rapid prototyping of possible alternative solutions in VR. This VR video has been used to support evaluation of solutions and fed into the training of the operators.

The use of 3D rapid prototyping at design stage was also deployed to support the identification and correction of critical Human factors issues that can improve the quality of the final products and prevent /limit impact of human error on it.

Another example on the operations side is the improved tools and capacity for Training. The analysis of critical activities and human barriers provides input to the training needs analysis whilst the prototyping tool for workplace design can be used for practicing the real job in a safe work environment using VR technologies.

In this effort to cover all aspect of normal operations we also now support the modelling of dynamic workflows considering staff responsibilities for specific task or projects, and embedding the required issue in relation to compliance with rules and regulations

TOSCA also designed an application called the Risk register as a prototype and a set of useful requirements for the establishment of a central tool for holding all identified and analysed risks. This will allow priorities to be established in terms of risk management, and help management to identify and track the mitigations to be put in place to manage the businesses' risks. Over time, as more information is logged in the risk register, trends can be analysed to help identify the company's safety performance. The risk register can also be used as a database, which can be mined to identify effective mitigations for known risks and as a piece of a common risk picture of the company shared among all the stakeholders. All the above tools and methods were born from embedding systemic human factors engineering methods as part of more traditional risk assessment and management approaches (TOSCA 2015)

References

TOSCA 2015 www.tosca-project.eu last accessed 10th of June 2015.