Irish Ergonomics Society

Irish Ergonomics Review

2011

Proceedings of the Irish Ergonomics Society Annual Conference 2011

Edited by

Leonard W. O'Sullivan and Chiara Leva

ISSN 1649-210

Table of Contents

ERGONOMICS FACTORS IN AGRICULTURAL HEALTH AND SAFETY

David H O'Neill

Dave O'Neill Associates
Clophill,
Bedford, UK

Abstract
Data on agricultural accidents in the EU have been reviewed with a focus on the statistics for Ireland and the UK. The reporting of fatal accidents is more reliable than that for the non-fatal for the reasons given. The greatest risks are associated with vehicles, falling objects and falls and these should be managed by the development and introduction of safer systems of work. The review has revealed that such interventions should be targeted at older workers and those who are self-employed. People who are in both categories should receive particular attention and, in Ireland, extra emphasis should be placed on older workers handling livestock.

Introduction

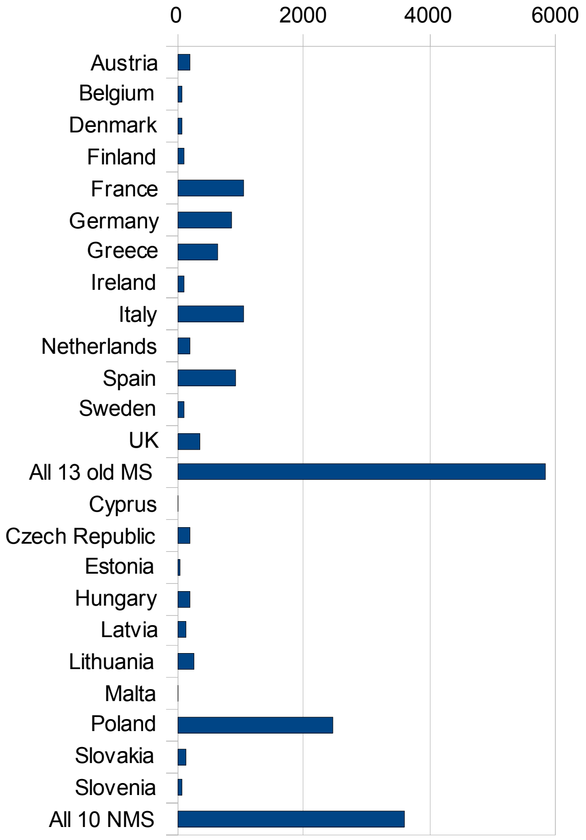
The number of people employed in agriculture in Europe has been declining steadily over the last 60 years, with the decline in the last 15 years being about 20%. Despite this, agricultural production and productivity have been rising because of the use of progressively bigger and more powerful machinery. One of the consequences is that the numbers of serious and fatal accidents and injuries have not declined at the same rate as employment. This paper will present some recent accident statistics, focusing on fatal accidents, for Ireland and the United Kingdom and examine the ergonomics implications. Where possible, the relevant European statistics are given to provide a broader perspective; in these cases the data are for "A" in the NACE classifications[1].

Over 10 million people work in agriculture in Europe. Although the numbers working, and the numbers of farms are reducing, it remains a vitally important business. Agricultural statistics, particularly regarding employment and accidents, vary quite widely between all the 27 EU Member States. Figures 1a and 1b show the numbers employed and the percentages of total employment, for 23 Member States (MS). These include 13 old MS (OMS) and 10 new MS (NMS) as indicated on the axes of the histograms. Of the original 15 OMS, there are no reliable data for Luxembourg and Portugal so they have been excluded.

Figs 1a and 1b do reveal differences between the OMS and NMS, mainly in the percentages of the total labour force engaged in agriculture. On average in the OMS, 3.6% of employment is in agriculture whereas in the NMS it is 12.4%. In Ireland there are fewer people working in agriculture than in the UK but in Ireland they comprise a greater proportion of the working population. Of the 13 OMS, Ireland has the second highest proportion working in agriculture, exceeded only by Greece.

1 see *ec.europa.eu/environment/emas/pdf/general/nacecodes_en.pdf*

The main sources of information in this brief paper are the relevant agencies of the European Commission and the Government sponsored Health and Safety bodies of Ireland and the UK (HSA and HSE). There is a further source of information which should be of interest to those concerned with health and safety on Irish farms – a doctoral thesis (University College Dublin) by Anne Finnegan[2].



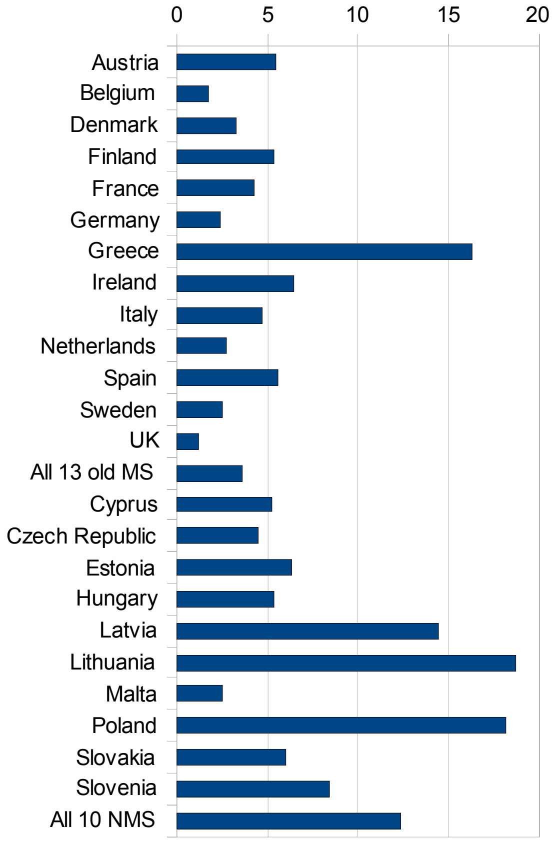


Fig 1a Numbers working in agriculture

Fig 1b Agricultural work as % of total

Accident statistics

Fatal accidents

The incidence rates for Ireland, the UK and the OMS (EU15) are given in Table 1 for the years where data could be found. The data sources are the Irish HSA, www.hsa.ie/eng/Your_Industry/Agriculture/ ,the UK HSE, www.hse.gov.uk/statistics/industry/agriculture/injuries.htm and the European Commission (2009).

2 Finnegan, A. 2007. An Examination of the Status of Health and Safety on Irish Farms. UC Dublin

Table 1 Incidence rates of fatal accidents in agriculture (per 100,000 people employed)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Ireland					13.3	14.7	15.5	17.7	15.5	13.4
UK	7.7	10.3	9.2	9.2	10.5	9.7	7.9	7.9	10.5	5.1
EU15	12.6	11.8	11.2	10.2	12	10.1				

According to EU-OSHA, the standardised fatal accident rate for agriculture of 12.6 per 100,000 workers for the OMS in the year 2000 does not compare favourably with that of 5.5 for all sectors combined[3]. For Ireland the average incidence rate or the period 1998 to 2007 was 15.0, which does not compare favourably with 2.0 across all sectors and, similarly, for the UK in the period 2004 to 2009, an incidence rate of 8.6 does not compare favourably with 0.7 across all sectors. The actual numbers of fatalities for Ireland and the UK are shown in Table 2, which includes data up until 2010.

Table 2 Total numbers of fatal accidents in agriculture

	2004	2005	2006	2007	2008	2009	2010
Ireland	13	18	18	11	21	11	25
UK	44	45	42	43	48	30	45

Both countries show a sharp increase for 2010 over 2009, but for the UK, this looks like a return to typical values, whilst for Ireland, at 53% of all workplace deaths it is a worrying increase over typical values.

Serious non-fatal injuries

The incidence rates for accidents and injuries causing more than 3 days absence from work for Ireland, the UK and the OMS (EU15) are given in Table 3 for the years where data could be found. The data sources are the Irish HSA, www.hsa.ie/eng/Your_Industry/Agriculture/ ,the UK HSE, www.hse.gov.uk/statistics/industry/agriculture/injuries.htm and the European Commission (2009)[4]. Table 3 also shows the EU15 data across all sectors

Table 3 Incidence rates of serious accidents in agriculture (per 100,000 people employed)

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Ireland					6500	8300	7800	7300	7500	7300
UK	490	610	540	400	340	330	320	410	440	450
EU15	6625	6159	5193	5048	5068	4560				
EU15 all sectors	4016	3841	3529	3329	3176	3098				

According to EU-OSHA (www.osha.europa.eu/en/sector/agriculture) the rate of around 6000 (per 100 000 workers) for accidents in agriculture with more than 3 days absence is amongst the highest rates for any industry. The data for Ireland are derived from all non-fatal injuries and are not limited to those injuries causing absences of more than 3 days. Therefore, compared with the EU15, these incidence rates would be expected to be relatively high. On the other hand, the rates for the UK seem remarkably low, but, according to the HSE these

3 EU-OSHA, 2001. Preventing accidents at work. EC Publications, Luxembourg. ISSN 1608-4144

4 EC, 2009. Causes and circumstances of accidents at work in the EU. ISBN 978-92-79-11806-7

rates are around then highest in any industry sector. These discrepancies in the data can may arise for several reasons:- a) the Agencies not following the same reporting criteria, b) under-reporting of accidents, which is known to occur for rural workplaces / occupations, c) the greater level of self-employment in agriculture compared to the other sectors, d) the informal nature of some agricultural work, e) the author's interpretation of the Agencies' statistics. In Ireland, the HSA[5] have stated there is a gross under-reporting of farm accidents by farmers, despite a legal obligation to do so.

In view of this, it would be preferable for any interventions or bases for further action to based on analyses of fatal accident statistics.

Nature of accidents and victims

Causes of accidents
The categories for reporting the causes of accidents in Ireland and the UK have a considerable degree of overlap, whereas the causes reported in the EU statistics are based on a somewhat different approach (for example see European Commission, 2009).

Table 4 shows the underlying causes of farm deaths and their proportions in Ireland and the UK over the ten-year period 2000 to 2009.

Table 4 Causes of agricultural fatalities

Code no.	Cause	Ireland (%)	UK (%)
1	Vehicles	49	25
2	Livestock	15	9
3	Drowning / asphyxiation	14	10
4	Falls	10	16
5	Falling objects	5	16
6	Wood related	4	inc with 5 above
7	Contact with electricity	2	3
8	Fire	1	--
9	Contact with machinery	inc with 1 above	9
10	Trapped by something collapsing / overturning	inc with 1 & 4 above	7
11	Other	--	5
	Total	100	100

Age of victims
Table 5 shows the age ranges of the victims of the fatal accidents over the ten-year period 1996 to 2005. The age ranges reported by the HSA and the HSE are not the same so some have been combined for a clearer comparison.

5 HSA. Farm Safety Plan 2003 – 07

Table 5 Age ranges of agricultural fatalities, 1996 to 2005

<table>
<tr><th>Age range</th><th>Ireland (%)</th><th>UK (%)</th></tr>
<tr><td>< 5</td><td>8.3</td><td rowspan="3">0.01</td></tr>
<tr><td>6 to 10</td><td>6.6</td></tr>
<tr><td>11 to 16</td><td>6.1</td></tr>
<tr><td>17 to 34</td><td rowspan="3">47</td><td>23</td></tr>
<tr><td>35 to 54</td><td>35</td></tr>
<tr><td>55 to 64</td><td>21</td></tr>
<tr><td>> 65</td><td>32</td><td>21</td></tr>
<tr><td>Total</td><td>100</td><td>100</td></tr>
</table>

In both countries it may come as surprise to find so many fatalities of workers over the age of 65 but that is the nature of farming. The causes of death for the older victims are distributed much the same as for all victims, as shown in Table 4, although in Ireland there seems to be a greater vulnerability to falls and livestock deaths with less of a threat from drowning.

For the 31 countries covered in the Fourth European Working Conditions Survey (European Foundation for the Improvement of Living and Working Conditions, 2007)[6] the age distribution across 13 industrial sectors showed agriculture to have a highest percentage of workers aged 55 years or older. This was 22% with the next highest being 15% for "Other Services".

Table 6 shows the age ranges of agricultural workers and those of workers across all sectors in the UK. This indicates specifically that the workforce is generally older in the agricultural sector than elsewhere.

Table 6 Age distribution (%) in UK agriculture and all sectors

	up to 24 yrs	25 – 39 yrs	40 – 54 yrs	55 yrs or more
Agriculture	8.9	27.3 *	38.5	25.2 *
All sectors	11.8	36.8	38.6	12.7

indicates $p<0.001$

Table 7 shows the age distribution for family farms in Ireland. These are reproduced from the HSA Farm Safety Plan 2003 – 07, which quotes CSO, 2000, as the source. As the intervals are different from those used in Table 6, direct comparison is limited but it is clear that an even greater proportion of farmers in Ireland are 55 or over than in the UK (nearly 40% compared with 25%).

6 European Foundation, 2007. Quality report of the 4th European Working Conditions Survey. Dublin EF/06/98/EN

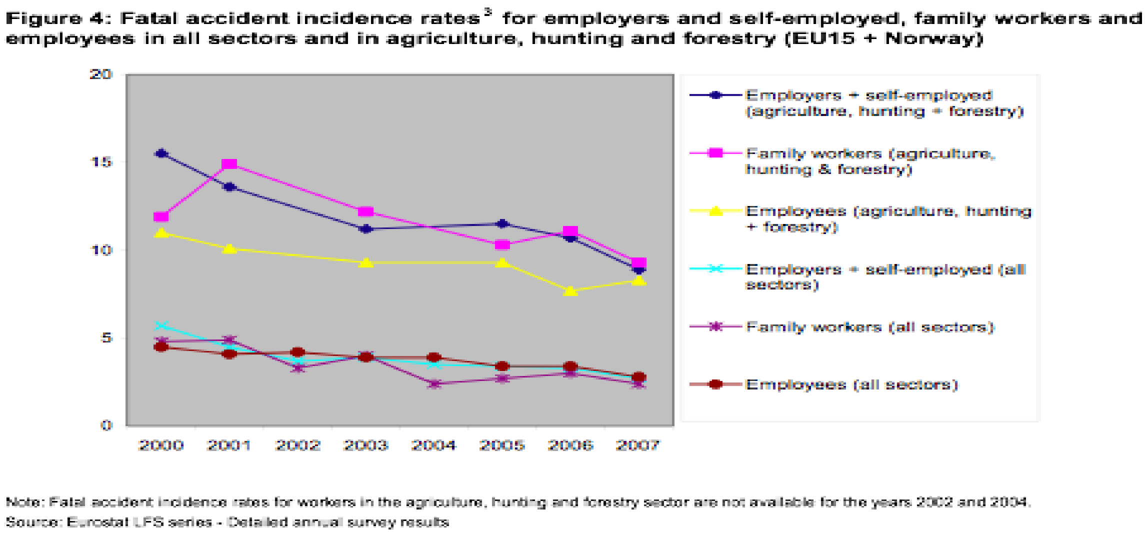
Table 7 Distribution of family farms in Ireland by age

Age range	Number	%
< 35 yrs	18400	13
35 to 44 yrs	30800	21.8
45 to 54 yrs	36300	25.7
55 to 64 yrs	27800	19.7
> 64 yrs	28000	19.8
Total	141300	100

Employment classification

In the EC database, fatal accident rates are similar for employees and the self-employed across all sectors except one. That exception is the agricultural sector (including hunting and forestry). According to EU-OSHA (2010)[7], in the 27 EU Member States in 2009, there were approximately 4.5 million self-employed people working in agriculture, more than in any other single sector. This is also confirmed by a report from the European Foundation for Living and Working Conditions[8] which states that in agriculture self-employment is prevalent (59%) compared to an average of 17% across all sectors

Picture 1 below, which is reproduced from EU-OSHA (2010), shows fatal accident incidence rates for employees and the self-employed, as indicated in the caption.



Picture 1 Fatality rates according to employment status (EU-OSHA, 2010)

The fatal accident statistics for UK agriculture for the years 2004 to 2009, according to employment status are shown in Table 8. (The totals for 2005-06 and 2006-07 include 3 and 4 children respectively.) Similar data for Ireland do not seem so readily available but the has in its Farm Safety Plan 2003 – 07 commented that, in 2000, there were 142,900 landholders and 141,500 farms, implying that many farms must be operated as one-person units. Farmers

7 EU-OSHA, 2010. A review of methods used across Europe to estimate work-related accidents and illnesses among the self-employed. ISSN 1831-9343 (TE-32-10-430-EN-N)

8 Eurofound, 2008. Factsheet, Agriculture and Fishing. EF/08/14/EN1.

were, therefore, carrying out many activities alone and likely to be under increased pressure to get time-critical work completed. This is likely to have an adverse effect on safety.

Table 8 Fatal accidents and employment status in UK agriculture 2004-09

	2004-05	2005-06	2006-07	2007-08	2008-09
employed	16	11	14	22	10
self-employed	26	23	22	24	16
non-employed	3	8	7	2	4
Total	45	42	43	48	30

The fatalities in the UK according to both age and status for the period 1998 to 2008 are given in Table 9. The totals increase steadily as the age ranges become higher but this is influenced mainly by the self-employed, as would be expected because very few employees would be kept on after the usual (male) retirement age of 65.

Table 9 UK fatalities according to age and employment status (1998 - 2008)

Age	Employees	Self-employed	Total
< 16	1	0	1
16 - 19	9	3	12
20 - 24	16	3	19
25 - 34	28	23	51
35 - 44	24	44	68
45 - 54	29	42	71
55 - 64	26	59	85
> 64	12	79	91
Age not known	2	5	7
TOTAL	147	258	405

Ergonomics implications

It is a serious concern for the Safety and Health Agencies around Europe that the incidence of accidents and injuries, including fatalities, is so resistant to reduction. As shown by Table 1, agriculture is one of the most dangerous sectors in which to work and does not seem to be getting any safer. The main underlying causes of accidents are similar in Ireland and the UK and can be addressed through the development and/or application of safe systems of work (SSW), as they all arise in fairly complex working environments.

The statistics on age and employment classification show that older workers and the self-employed suffer a disproportionate number of accidents compared with workers in other sectors. Whilst they are all affected by the hazards implicit in Table 4, in Ireland specifically, it would seem that safer livestock handling by older workers should be given extra emphasis. According to Finnegan's thesis, activities associated with livestock dairying, rearing, husbandry) present a greater risk than vehicles (see Table 4) but this may be an outcome of her method of classification.

The agricultural profession operates under a unique set of circumstances (which are well understood) but they are likely to require a unique set of interventions to manage the risks, especially those which threaten older or self-employed workers, and, of course those in both categories. The challenge for the ergonomist or health and safety professional is to devise safe systems of work that are found acceptable in these unique sets of circumstances.

Disclaimer

Any errors or omissions in analysis and interpretation of the data are entirely the responsibility of the author.

THE CHALLENGES OF ERGONOMICS PROGRAMME IMPLEMENTATION ACROSS MULTIPLE COUNTRIES – ONE DEPARTMENT'S STORY

Fiona Moclair

Intel Ireland,
Collinstown Industrial Park,
Leixlip,
Co. Kildare

Abstract

This paper outlines the unique challenges posed by the implementation and proliferation of Intel's ergonomics programme across one department. The department, Corporate Services EMEA has personnel based throughout the EMEA region. Due to the multiple sites involved, novel approaches were required with respect to the various elements of implementing the ergonomics programme. This paper focuses on ergonomics training, ergonomics assessments and equipment such as mice, office furniture etc. The paper finishes illustrating the ergonomics topics that the department is currently engaged in.

Introduction

Intel Corporation is commonly known as a semiconductor manufacturing company. However, there are other aspects about Intel that are not commonly known. It is one of the largest venture capital companies in the world and after Microsoft; it is the second largest software company in the world.

Intel's ergonomics programme was established back in 1991 with the hiring of a Corporate Ergonomist. From then, the programme was proliferated across the company on a global basis. Amongst other elements, the programme emphasises ergonomics training, assessments and equipment focused on the end user.

Corporate Services EMEA

The department of interest, adapting and implementing the ergonomics programme is Corporate Services EMEA (CS EMEA). It consists of 175 employees based throughout the EMEA region at multiple sites. It is essentially a support organisation for Research & Development, Logistics, Sales & Marketing and centralised functions such as Finance, Human Resources…basically all of the organisation that are not directly involved in manufacturing.

The department is predominantly office based with some technicians at the larger sites.
CS EMEA provides a myriad of services including:

- Physical infrastructure for Intel employees
- Facilities services
- Food services
- Security
- Environmental, Health and Safety
- Construction
- Landscaping
- Company car fleet

Ergonomics Programme Challenges

The department consists of 175 employees based across multiple EMEA sites. It has a significant geographic dispersion with varying time zones, six time zones in total. In some of the countries in which the department operates from, ergonomics is not known. Although English is the day to day operational language, for the vast majority of the sites, English is not the mother tongue. Local expertise was not available at all sites. In addition, regulatory requirements also varied to those within the European Union and those outside of it. Given the number of sites where CS EMEA personnel are based, it wasn't feasible to travel to each site. This resulted in the need to utilise remote means of implementing the various ergonomics programme elements.

Ergonomics Training

Intel developed its own in house multimedia Office Safety and Ergonomics course. It was decided that all of the CS EMEA department would take this course as a first step in increasing their overall knowledge of ergonomics. Having the course in a "Web Based Training" (WBT) format provided an interactive multimedia experience for the folks taking the course. It also allowed employees to take the course at a time suitable for them. Unlike instructor led courses, it avoided the need to solicit volunteers to teach the course and undertake train the trainers for these volunteers.

Ergonomics Assessments

The department had two methodologies that it could use to conduct ergonomics assessments. Where the resources were available, local ergonomics assessors were utilised to conduct 1:1 ergonomics assessments. A remote ergonomics assessment was conducted for employees that didn't have the opportunity to have a 1:1 ergonomics assessment.
This consisted of the employee getting digital photographs taken of themselves while they were working at their PC. The photographs were sent to their designated ergonomist. In a scheduled meeting, the employee met with their ergonomist to review the photographs over a conference call and live meeting. Recommendations were co-developed during the meeting to improve the employee's physical environment.

Ergonomics Equipment

Ergonomics equipment falls under two distinct categories:

- Personal equipment such as PC mice, laptop bags, monitor blocks etc.
- Infrastructural equipment such as office task chairs, desks etc.

Based upon the ergonomics assessment undertaken, each employee would have had equipment recommended and this would have been procured for them by their department. All items chosen are selected from an EMEA “approved” selection of accessories.

Current Ergonomics Focus Areas

Building on the ergonomics initiatives done to date, the department’s focus has expanded to the application of ergonomics in “non office” environments. This covers working in the following areas:

- Conference rooms
- Cafeterias
- Home offices
- Travelling
- The use of mobile computing devices such as smartphones, tablets etc.

SAFETY AND/OR HAZARD NEAR MISS REPORTING IN AN INTERNATIONAL ENERGY COMPANY

P. Clancy[1], M.C. Leva[2], V. Hrymak[3], Michael Sherlock[1],

[1]*Asset Management Services, ESBI Ireland*
[2]*School of Psychology Trinity College Dublin Ireland,*
[3]*School of Food Science and Environmental Health Dublin Institute of Technology*

Abstract

This paper presents the preliminary progress of an industry driven programme to improve the data monitoring of safety/hazard near miss reporting from front line staff of a branch of a multinational energy supply company in Ireland. The paper discusses the main factors that emerged as possible causes for underreporting and the course of action selected for addressing them. The initiative, which is only in operation for 4 months has already led to an increase in reporting of "near misses" by a factor of nine. Furthermore, the level and detail of the reports is far greater than had previously being received.

Introduction

It is recognised as Best Practice that collating and monitoring of data on safety incidents or reported near misses would lead to better learning and indeed the avoidance of accidents in the future. (Jones et al 1999). Various studies have showed the relationship between near miss incidents and actual accidents pointing out that reducing the number of near misses that occur will very likely reduce the likelihood of full accidents, which in turn would lead to less severe human, economic and environmental impact., (Bird & Germain, 1966; Heinrich, 1980; Tye, 1976).

Within this paper, the experience of the safety advisor within a division of a larger multinational energy supply company was able to identify from report statistics supplied by the Head Health and Safety Office that near miss data from field staff in his division as a whole was not being reported. This data has the potential to identify latent hazards in plant, equipment, procedures and in design of high voltage (HV) equipment that may otherwise go unnoticed.

The division in question is the Asset Management Service, AMS, within the company, which is responsible for providing a full range of commissioning services on new and maintained HV plant & equipment. AMS also ensures the correct operation of protection schemes on the Transmission systems.

During 2009 this section of the company only had only reported 2 near misses for events of trivial importance related to offices and nothing referring to site operations. The 80 staff in the division spend the majority of their time in the field on site works. The fact that there was no reporting from the field was of concern to the senior management team in AMS.

The Issue of Underreporting

Various studies on organizational-level under-reporting linked the issue to multiple factors. Typical issues would be the general safety climate, the specific industrial, sector, the company size and the perceived lack of management engagement (Leigh et al., 2004; Oleinick et al., 1995, Daniels and Marlow, 2005, Clarke, 1998, Probst et al., 2008; Zohar, 2003) at individual-level under-reporting has been ascribed to factors such as fear of reprisals, loss of benefits or a fatalistic attitude that injuries are a fact of life in certain lines of work (Webb et al., 1989;Pransky et al., 1999; Sinclair and Tetrick, 2004, (Pransky et al., 1999).

From a previous study performed in Ireland in the construction sector (McDonald N and Hrymak V. 2002) it appears also that the presence of a safety representative on site shows a very strong relationship with hazards reporting and safety compliance. The report of the study states that "safety representatives influence safety compliance not only through their influence on the response to audits and hazards but also through other means. Thus they encourage the reporting of hazards and help ensure that these reports lead to better safety compliance on site. Their presence also makes it significantly less likely that workers will continue to work in hazardous situations".

In the context of the present study the positive effects of the safety representative on site were reinforced thanks to the presence of the specific organizational role played by Safety Advisor within the specific division.
The Safety Advisor in this case was an interface between the central Health and Safety Department and Asset Management Services and He was tasked to take care of the division specific safety issues involved in the day to day operations.

The Safety Advisor was able to work on site with the staff and perform informal interviews in order to try and identify the main issues for this lack of reporting. Through this process three factors emerged as possible causes for underreporting:

1. The current definition of "near miss reporting " and indeed the actual safety training received by the staff were confusing.
2. The actual reporting framework was received as extra paper work to be sent to the immediate supervisor in a very formal process.
3. Poor feedback on reported problems.

Overall the "near misses" reporting process was seen as something the staff was told to do rather than something they should be doing for their own benefit. As a result it was perceived as "an extra task not a value".

A simple plan for action

In identifying the best course of action to take to try and improve reporting of the following elements were taken into account:

A. A dedicated safety advisor for the Asset Management section to bridge the gap between the Safety Management System and day to day operations.
B. a different definition of near misses that would highlight the relevance in respect to the everyday operations and a proper communication of it to the workers

C. A reporting form more closely related to forms currently part of day to day usage
D. A feedback mechanism to ensure the benefits of reporting in terms of follow up would reach the front line staff in charge of reporting.

The Role of the Safety Advisor

As already pointed out the operational safety advisor for the Asset Management Services (AMS) division of the company is a specific recognised organizational role established to take proper action for the safety issues of the division of a specific technical nature rather than the management of occupational health on site, which was still dealt by the Central Health and Safety Office. The Safety Advisor in fact had to also develop personal technical competencies for AMS in parallel with the safety role.; which in turn enable him/her to work side by side with the rest of the technical staff of the division.

The availability of a safety advisor close at hand similar to a safety representative but with the managerial role to enforce and follow up on issues raised on the field was highlighted as a strong guarantee towards the achievement of a better promotion of a reporting culture and a closer feedback to the front line staff, in line with the findings of the HSA research report for the construction sector (McDonald and Hrymak 2002).

Near Misses, definitions and communications

The definition of near misses previously provided by the Central Health And Safety Office of the company stated that:

'A near miss is an incident where personal injury was narrowly avoided or where damage to property-only occurred. A good catch is an unsafe condition/act, which if left unaddressed could result in an injury. Such incidents may be early warning signs of hazards that could eventually result in serious consequences. By reporting such incidents you will help make the workplace safer for yourself, your colleagues and visitors. Remember, what is a near miss today could result in an accident tomorrow.'

The staff understanding of the above definition also reinforced by the type of information provided by newsletters and the periodic training promoted by the Central office was that a near miss belonged only to realm of occupational health on site and did not apply to specific technical issues related to operations. The link between the day to day anomalies in the field and the ability to make work practices more efficient, of a higher quality and safer was not being recognised.

The Safety Advisor was able to introduce among his fellow workers an alternative definition and to promulgate it through a specific meeting.

'An opportunity to improve safety, health, environmental and quality practice based on a condition, an incident or an observation with minor outcomes but with the potential for more serious consequence. '

The workers were made aware that the consequences proposed in the definition can include but were not limited to the following:

- Property damage
- Damage to the environment
- Business interruption
- Deviations for example from the work instruction or procedure

- Potential or actual injury to staff

The definition above was presented and discussed with the personnel by the safety Advisor in the following ways:

i. Organizing a meeting with the engineering manager where it was agreed to produce a document outlining the benefits of an integrated approach to the management of quality, safety, health and environmental issues.
ii. Providing a presentation during a periodic team briefing meeting in the AMS section where the alternative definition was discussed and amended
iii. Sending a communication through email to all AMS staff
iv. Reiterating on the presentation regarding the near miss management approach at Specialist Team meeting and subsequent team meetings attended
v. Promoting the idea also informally on site

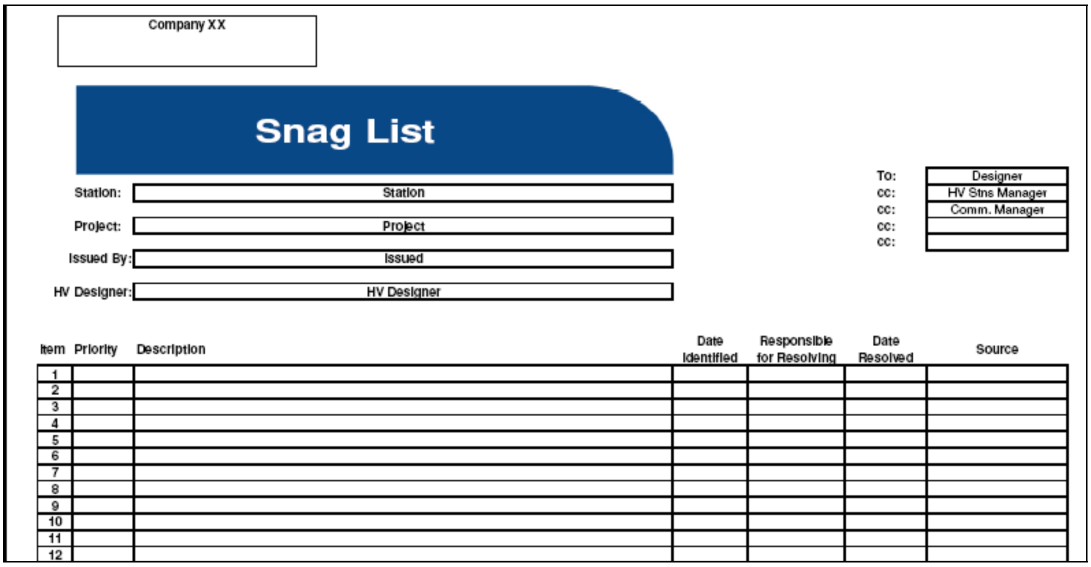
The initial feedback was that the new definition was accepted as a better fit to the working environment in AMS since it gave the opportunity to report or capture technical deviations or observations that commissioning and maintenance staff can encounter in their daily tasks.

A new reporting framework: making better use of what is already there

The new reporting framework for the initiative was introduced as an informal process. The future direction of it to be embedded in the commissioning checklists that are already used as part of the sign off for the operators working on site.

Currently commissioning based reporting has two elements. The first is the on-site Snag List Form(see figure 1) and the other is the Project Follow-Up Reporting form. The snag list is to be given to the person on site responsible for correcting it and then entered into a so called "SharePoint folder" where the design team can monitor it and make the any design changes.

The Project Follow-Up Report form is given to the people identified as responsible for solving the issues and also sent electronically to the designer and manager of substation design. This form should also be copied to the SharePoint folder where the actions can be monitored. The deviations recorded on the snag list However or the follow up report forms are not currently considered as possible elements of the miss reports statistics at central level.



Company XX

Snag List

Station: Station

Project: Project

Issued By: Issued

HV Designer: HV Designer

To:	Designer
cc:	HV Stns Manager
cc:	Comm. Manager
cc:	
cc:	

Item	Priority	Description	Date Identified	Responsible for Resolving	Date Resolved	Source
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

Figure 1: company existing Snag list for commissioning operations

Further a Database has been introduced to collect and monitor the commissioning checklist as records of the operations completed.
The database has a function that provides report templates for commissioning and "condition based assessment" of assets and records the results. These templates are designed by AMS staff. The templates also provide the option of recording deviations in an 'additional comments' field and to import or export documents/files/jpeg/PDF. Reports can also be generated for any specific asset.
The safety advisor prompted the workers to start using that part of the current reporting forms for forwarding information on near misses via mail to him and whenever possible attach pictures of the possible event being reported. Figure 2 reports the example of a picture attached to one of the reports ("Failure of 110 kV cable sealing end").

Figure 2: Example of a picture taken by one of the worker and sent as part of a near miss report on the failure of 110 kV cable sealing end

The advantage of using existing tools for reporting is that the use of ad hoc extra forms may fail to provide a real-time picture of routine operations supporting performance management and predictive risk management. Furthermore the use of many discrete tools implies that much valuable data gathered about the operation are stored and analyzed in different formats and by different and often disjointed departments. This makes it difficult to obtain an integrated risk registry (Leva et al 2010), while the effort of integrating existing data collection tools can be a much more practical way to operate the data monitoring reducing the paperwork.

Another issue existing with the previous reporting system is the fact that the health and safety office at central level presents a classifications system for near misses with categories that are fairly generic and therefore are not able to really direct possible improvements and follow up initiative on specific technical areas in a meaningful way. An example of this is observable from figure 3 reporting the headings under which near misses are currently categorized. It is clear that a category named "electrical" is far too generic to be able to provide any clear indication for a division in charge of commissioning HV equipment and installations. The proposed enhancement of the classification introduced by the imitative would only require to distinguish electrical faults according to the type of equipment they refer to (e.g. Neutral Earth Switch, Cable Sealing Ends Links , Busbars, G.I.S Switchgear , HV Lightning Arrestor, HV Transformer Bushing, Oil Filled Circuit Breaker , Capacitor Bank etc..). This further distinction would enable also to classify possible troubleshooting adopted for recurring faults.

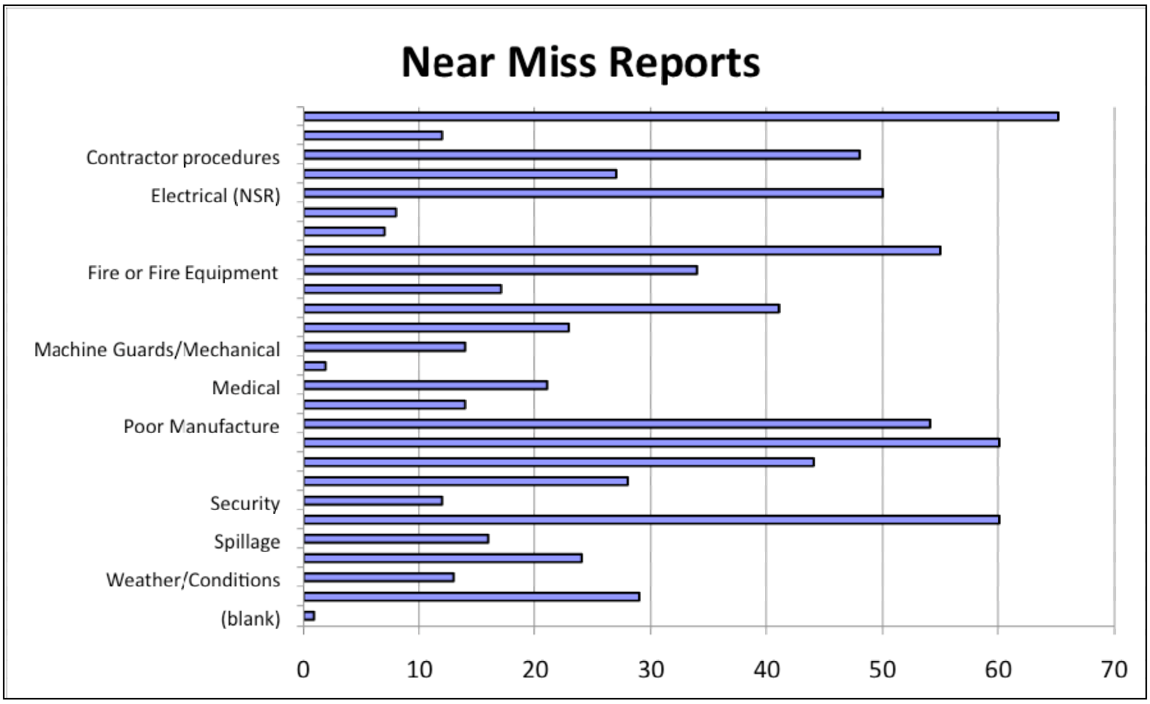


Figure 3: Existing near miss categories for event reported across the entire company in 2009

The feedback mechanism

As already pointed out the main purpose of reporting near misses is the possibility of using the resulting data to initiate improvements and possible interventions able to prevent more serious events and accidents. Further the feedback to the reporters on the follow up initiated thanks to their reporting is an important motivational factor. Therefore the safety advisor had to take care that after each report a communication about the status of the analysis and the possible initiated action would reach the front line staff in charge of reporting.

Since the beginning of the initiative 29 near misses have been reported in the past 12 months and the follow up of 62% of them was already completed and communicated to the report initiators.

Preliminary Results

The initiative in the first 12 months was already able to increase by 14 times the amount of reports that were previously obtained within a year, further the level of technical details acquired and their relevance is much more meaningful and in depth in comparison to what collected previously.

Table 1: Initial outline of results of the Initiative

Near Miss Reports	Amount	Location
Centrally collected Events in 2009	768	Site / Office
Events collected at the division level of Engineering Solutions 2009	60	100% of them are related to the Office
Near Misses collected for AMS in 2009	2	100% of them are related to the Office
Near Misses collected for AMS in 12 months after new definition was introduced (12 months approx.)	29	90% of them are related to Site

Conclusions

Data collection programs such as these provide a real-time review of current safety issues in the operations departments. Real-time data review facilitates the identification of areas where modifications to working practices, equipment, training programs or standard operating procedures might be appropriate. Such modification might reduce costs as it improves the availability of equipment and prevent the occurrence of future safety events (incidents or accidents) as well. This seems to be a very proactive way of managing safety with very positive implications for day to day operations efficiency as well. The key to success is arriving at the desired cultural climate as a result of the system changes introduced. This is why a careful understanding of people dynamics is not to be underestimated. In the present experience the role of a safety advisor close at hand on site with the capacity to follow up on issues raised on the field was highlighted as a very important factor towards the achievement of a better promotion of a reporting culture and a closer feedback to the front line staff.

References

Clarke, S., 1998. Organizational factors affecting the incident reporting of train drivers.Work & Stress 12, 6–16.

Bird, F. E. & Germain, G. L. (1966). Damage Control. New York: American Management Assoc. Inc.

Daniels, C., Marlow, P., 2005. Literature Reviewon the Reporting ofWorkplace Injury Trends (HSL/2005/36). Health and Safety Laboratory, Harpur Hill, UK.

Heinrich, H. W. (1980). Industrial Accident Prevention: A Safety Management Approach, (5th ed.). ISBN 0 07 028061 4.

Jones S., Kirchsteiger C., Bjerke W., The importance of near miss reporting to further improve safety performance. Journal of Loss Prevention in the Process Industries 12 (1999) 59–67

Leigh, J.P., Marcin, J.P., Miller, T.R., 2004. An estimate of the U.S. government's undercount of non-fatal occupational injuries. Journal of Occupational and Environmental Medicine 46, 10–18

Leva M.C Cahill J Kay A Losa G. Mc Donald N. 2010 The Advancement of a New Human Factors Report – 'The Unique Report' - Facilitating Flight Crew Auditing of Performance/Operations, as part of an Airline's Safety Management System. Ergonomics Feb;53(2):164-83.

McDonald N and Hrymak V. Safety Behaviour in the construction Sector. HAS Research Report 2002.

Oleinick, A., Gluck, J.V., Guire, K.E., 1995. Establishment size and risk of occupational injury. American Journal of Industrial Medicine 28, 1–21.

Pransky, G., Snyder, T., Dembe, A., Himmelstein, J., 1999. Under-reporting of workrelated disorders in the workplace: a case study and review of the literature. Ergonomics 42, 171–182.

Probst T.M. Estrada A. X., 2010 Accident under-reporting among employees: Testing the moderating influence of psychological safety climate and supervisor enforcement of safety practices. Accident Analysis and Prevention 42 (2010) 1438–1444

Probst, T.M., Brubaker, T.L., Barsotti, A., 2008. Organizational under-reporting of injury rates: an examination of the moderating effect of organizational safety climate. Journal of Applied Psychology 93 (5), 1147–1154.

Sinclair, R.R., Tetrick, L.E., 2004. Pay and benefits: the role of compensation systems in workplace safety. In: Barling, J., Frone, M. (Eds.), Psychology of Workplace Safety. American Psychological Association,Washington, DC, pp. 181–201.

Tye, J. (1976). Accident Ratio Study, 1974/75. London: British Safety Council.

Webb, G.R., Redman, S., Wilkinson, C., Sanson-Fisher, R.W., 1989. Filtering effects in reporting work injuries. Accident Analysis and Prevention 21, 115–123.

Zohar, D., 2003. The influence of leadership and climate on occupational health and safety. In: Hoffman, D.A., Tetrick, L.E. (Eds.), Health and Safety in Organizations: A Multilevel Perspective. Jossey-Bass, San Francisco, CA.

ERGONOMICS PAYS

Ernst A.P. Koningsveld

TNO, PO box 718,
2130 AS Hoofddorp,
The Netherlands

Abstract

With their work, ergonomists aim to provide safe and healthy environments and tasks, in which people can perform well over many years, up to a lifetime. Despite this undisputable goal, the impact of ergonomics efforts is not always as large as ergonomists and human factors specialists want. There is a trend to demonstrate the positive contributions of ergonomics interventions with economic evaluations. This paper starts with an overview of the origins of ergonomics. Those with a background in health prevention used to be reluctant from economic effects as a motivator for change. However, human performance based ergonomists, focused from the very start at the effects on business costs and performance. Based on a series of business cases, this paper illustrates that these two lines are not contradictory, in fact they go well hand in hand.

Introduction

This paper aims to illustrate the benefits of ergonomists' work. When looking at the origin of the profession, we see two basis backgrounds: one focuses on safety and health, the other on human performance. Over the years these two scopes seemed to be contradictory. Only when ergonomists started making economic analyses of their work, it proved that these lines are not contradictory. And even more important: many business cases demonstrate that ergonomics pays!

Ergonomics to promote safety and to protect health

For hundreds of years, workers' health protection and providing a safe working environment have been important. The relation between occupations and injuries was already documented centuries ago. Ramazinni (1633-1714) wrote about work-related complaints in the 1713 supplement to his 1700 publication, "De Morbis Artificum" (Diseases of Workers; Ramazzini, 1940).

Less known is Wojciech Jastrzebowski's work. In 1857 he created the word 'ergonomics' in a philosophical story, "Based upon the truths drawn from the Science of Nature" (Jastrzebowski, 1857). In the second half of the nineteenth century health impairment of the working population became an item of political interest. In the Netherlands, the protection of the health of children at work was the real start, with the Dutch Legislation on Child Labour in 1874. Legislation, inspections, interventions, trainings and education followed in many countries. Though much progress has been achieved in labour protection over more than a century, even today working is to often unsafe and overstraining, resulting in large numbers of early dropouts and health impairment. Next to a reduction of injuries and accidents, in the industrially developed countries the character of hazards moved from physical to socio-psychological.

The scientific knowledge of human's capacities extended largely. Limit values for actual workloads were defined and introduced at the level of legislation and standards, or as good-practice agreements between employers' organisations and unions. Solutions for hazardous workloads and unsafe conditions have been developed, tested and implemented. Nevertheless, even today, adverse working conditions do exist, even in the most industrially developed parts of the world and in high tech industries. Shareholders and management seem to misunderstand the long term benefits of safe and healthy work environments.

The consequences of poor working conditions are serious. Workers may suffer from health impairment and experience a too poor quality of life. Injuries reduce their power to earn money, ultimately affecting their independence in life and their families. Health effects amongst employees may confront the employer with high costs, though the direct financial effects differ widely over countries. In the Netherlands the employer is held responsible to pay at least 70% of wages during the first two years of sick leave (for whatever cause). In other countries this period is far less, and applies only cases of occupational accidents and injuries.

It is remarkable that the focus on productivity losses, quality problems, reduced motivation, or loss of unique expertise by occupational health impairment gets relatively little attention. In the western world, the added value of human activities at work is in the range of 2 to 5 times their salaries. So, it is easy to understand that suboptimal human performance affects the financial performance of organisations directly.

At the level of a sector, unhealthy or unsafe work results in unappealing jobs, making it hard to attract new staff. For instance, today, physically straining work is less attractive to youngsters than being a machine operator. The return on investments in vocational training can be limited.

At the societal level, the costs of poor working conditions are a real burden. Several European studies show that the societal costs are in the order of 2-3% of the Gross National Product (GNP) (e.g. Koningsveld, 2003). This Dutch study provides an overview of cost categories and their values, but the calculation does not include all effects. The costs of occupational accidents could not be assessed. The validation of the financial effects of poor working conditions on companies' performances turned out to be too difficult to calculate. So these costs are not included; some experts estimate that these may be in the range of another 2-3 per cent of the GNP. Others, including the International Labour Organisation (Forastieri, 1999) point at the importance of better working conditions for companies' performances. Many individual companies and organisations are cost driven; but according to Oxenburg (2004) and Marlow (2006), cost benefit considerations related to occupational health and safety are scarcely out of the egg.

It is not easy to express lost healthy lifetime and individual harm in financial terms. These factors are not included in the fore mentioned studies about the costs of poor working conditions. But today, these factors are becoming important matters in the scope of politicians, employers, workers, and their representatives.

Ergonomics to promote human performance

The development of human performance based ergonomics has its roots in studies on physical performance. Employers used to be highly interested in increasing human performance, and so in the maximum capacities of humans. Scientific Management, a method that improved worker efficiency by enhancing the efficiency of job processes, became popular. Frederick W. Taylor was a pioneer of this approach and evaluated jobs to determine the "One Best Way" they could be performed. At Bethlehem Steel, Taylor dramatically increased worker productivity when shovelling manually, by matching the shovel with the type of material that was being moved (ashes, coal or ore). Frank and Lillian Gilbreth made jobs more efficient and less fatiguing through time and motion analyses and standardizing tools, materials and the job process. By applying this approach, the number of motions in bricklaying reduced from 18 to 4.5, allowing bricklayers to increase their pace of laying bricks from 120 to 350 per hour (Ergoweb website).

From the nineteen-forties, ergonomists have been involved in cognitive, mental and organisational performances. As several ergonomics publications indicate, during World War II, more planes crashed on their way to the battlefields or on the return flight, than were shot by the enemy. A vision on 'human centred technology' was born, and has become important. Today, the fast development of new technologies makes this even more important.

Still a large gap exists between the designers and users of ICT. Designers show little understanding of the way humans deal with technology, or how they want or expect technology to work. As early as the nineteen-seventies, studies about operator's behaviour in control rooms and design recommendations were published by Edwards and Lees (1974). The European Coal and Steel Community put much effort in this field. Later, mistakes in control rooms resulted in serious accidents. Analyses of these accidents usually show human error failures that could have been solved by sound ergonomics system design. Examples of such accidents with proven human factors' misfits are: the Flixborough disaster (1974, an exploding chemical plant), the Three Miles Island nuclear incident (1979), and the Bhopal catastrophe (1984, release of toxic compound). Authors like Kirwan published several books on the matter (Kirwan, 1994), also focussing on the mental workload and the internal process representation of the human operator.

Other studies on human performances, based on mental processes, focus for instance on drivers' tasks in cars, trains, ships, and even in bicycle riding. Recent studies warn for the hazards of dual or multi-tasking, for instance mobile telephoning when driving.

Organisational design and management has become one of the large topics in ergonomics. Subtopics, like socio-technical design or participatory ergonomics get much attention at ergonomics conferences and in literature. Such subtopics are not completely free of hypes. Every decade new gurus or visions arise, many of which deal with human performance. Shapiro (1996) has written a rather critical review of work organisation design and management views, entitled Fad Surfing in the Boardroom: managing in the age of instant answers. Much organisational hype passed by, such as total quality management, systems re-engineering, Six Sigma, or learning organizations. Of course, each of these has its values. But Shapiro states that, at the end, the organisation's consultants are the only ones who benefit. Probably, this is because many of these approaches don't address the actual production tasks at the shop floor. And that is basically what ergonomists and human factors specialists do.

Ergonomics' benefits in practice

General benefits of ergonomics interventions

As stated, ergonomists should focus on the benefits of their work in terms of core business values. That means that the effects on the system's performance should be analysed. To prove effectiveness is a relatively new field of interest in the profession of ergonomics. Table 1 shows the effects that ergonomics interventions may have and that are core business values.
Not all of these can easily be presented as objectively assessed data. That may be a problem for ergonomists, but general management is used to work with a combination of quantitative and qualitative effects, and can very well base their decisions on those.

Costs and benefits evaluations
In the ergonomics society the demand for cost benefit analyses increased. Many ergonomists hope to be more convincing, if they could demonstrate the financial effects of their work. Some good case studies were done. However, for a long time the development of a generic model was considered too difficult. Oxenburg (2004) did great work in the field, but the need remained for an as simple as possible approach. In 2005 four approaches were presented and discussed at a WHO and NIOSH sponsored conference (Eijkemans & Fingerhut, 2005). None of these was considered to be the ultimate instrument. One was too time-consuming, another too complicated and a third one required much competence and expertise of the analyser.
Since then, TNO in The Netherlands did a lot on cost benefits and cost effectiveness of interventions. Several, more or less generic instruments were developed for sector organisations. For the greater part, these instruments use the same set of parameters (table 2). Basically, work as it is before the intervention, is compared to work after the changes. This can be done either as a virtual improvement before the implementation (by estimating the effects), or as an effect evaluation after implementation (using assessed data). The costs of interventions are expressed as investments and actual operational costs.

Business cases evaluated

The evaluation of financial effects and qualitative effects of projects can be a part of projects. Eighteen of such cases were reviewed by Koningsveld in 2008. The review intended to find out which factors were convincing for the decision to implement improvements in working conditions. The cases are diverse, ranging from ergonomically designed hand tools, via assembly work and an integral health program, to job enrichment. Seven of the eighteen cases show a return on investment (ROI) in less than 1 year, while two other have a ROI in a little more than one year (Table 3). Managers usually decide immediately to implement interventions with a ROI of 1 year or less. All the other cases are profitable within 3 years; many companies tend to decide positively about investments with a ROI of 3 years or less.

Table 1. Potential core business effects of ergonomics interventions (Koningsveld, 2005)

Increased productivity	**Lower operational costs**
More efficient movements	Fewer lost working days
Less fatigue	Fewer cases of disability
Better motivation	Easier and quicker return to work
Less personnel turnover	Fewer temporary workers
Fewer temporary workers	Lower costs to assist sick workers
Easier to assemble products	Fewer rejected products
Improved competitiveness	**Company's values/standards**
Improved production, product quality	We take health and safety seriously
Increased flexibility of production	We are proud of our workers
Improved workers' satisfaction	Improved safety: lower accident rate
Improved clients' satisfaction	Sustainable production, products
Higher reliability of delivery	Etc.
Better position on the labour market	

Table 2. Categories of costs and benefits

Operational parameters for core business (assessed before versus after implementation) - Man hours spent, costs per man hour - Costs of lost working days (corrected automatically for work related injuries) - Personnel turnover costs (corrected automatically for work related cases) - Costs of accidents, damage, juridical cases
Investments for ergonomics interventions - Consultancy costs to select and adapt an intervention; - Costs of time of own staff to select interventions; - Purchases: the direct expenses for hardware, software; - Installation costs, costs of introduction; - Initial costs, like training employees; temporary reduced production; - The amortisation term (1, 3, 5 or 10 years) and the interest rate applied; - Rest value: some investments may have a rest value after the term of amortisation; - Grants, subsidies minus the costs to acquire those.
Operational costs of interventions - maintenance costs; inspections; - energy and other consumables; - other structural expenses; e.g. travel costs, subscription costs, training of new staff.
Qualitative effects of the intervention – on the quality of products, services, processes; – on lead time, time to market; – on the labour market attractive power; – on flexibility of production, products, work processes; – on positions' fit for elderly workers, partly handicapped workers; – on the quality of work, on the level of work stress; – on employees' contentment score; – on clients' contentment score; – on innovative power of the organisation; – on complying with the core values of the organisation.

Table 3. Eighteen costs benefits cases compared (Koningsveld, 2008)

Intervention	Sector	Benefits per year/cost of intervention	ROI (years)	H&S benefits	Core business
ergonomic screwdrivers	many	299%	0,3	4%	96%
ergonomic workplaces	assembly grasmowers	64%	1,6	29%	71%
ergonomic workplaces	quality control microchips	66%	1,5	16%	84%
ergonomic workplaces	assembly emergency lights	1281%	0,1	54%	46%
smaller bricks	bricklaying melting ovens	171%	0,6	63%	37%
ergonomic cabin	public transport (streetcar)	103%	1,0	78%	22%
job enrichment	painters	108%	0,9	1%	99%
job enrichment	plasterers	42%	2,4	8%	92%
ergonomic tools	window pane mounting	95%	1,1	21%	79%
mechanical paving	road construction (paving)	57%	1,8	2%	98%
mechanical transport	road construction (paving)	154%	0,6	3%	97%
rolling carpet in van	parcel delivery	84%	1,2	5%	95%
ergon. vacuum cleaners	professional cleaning	211%	0,5	6%	94%
sit-stand office desk	office work	69%	1,5	32%	68%
integral health program	hospital	424%	0,2	52%	48%
safe road blocking	road construction	38%	2,6	0%	100%
patient handling lifts	care of handicaped	60%	1,7	3%	97%
wandering patients	care of demented elderly	99%	1,0	1%	99%

In the review, the benefits are divided into two categories: traditional occupational safety and health benefits (injury and accident prevention, lost working days, disability), versus 'core business values' (productivity, direct costs, extra output, failure costs, quality). The scope of core business values is in line with the 'human performance' indicators that are used in this paper, generally a combination of parameters. In the cited article however, the benefits for these parameters were first identified, then quantified where applicable, and finally added up. Despite the fact that almost all projects start from the OSH perspective, in all but one case, both core business and OSH benefits occur as a result of the intervention. However, the core business values of fourteen of the eighteen cases do exceed the OSH benefits. In ten of these, the core business benefits represent more than 90% of the total benefits. Only in two cases, the H&S-benefits evidently exceed the core business ones.

With the limitations that these cases are all performed by one group of ergonomists, the evaluation makes it evident that the prevention of unsafe working conditions and health impairment goes hand in hand with enhanced human performance. The review proves as well, that initiatives to prevent health hazards and to promote safety often result in better organisational performance. To conclude: ergonomics interventions do pay!

References

Edwards E, Lees FP 1974, *The human operator in process control*. Taylor & Francis Ltd., London.

Eijkemans G, Fingerhut M (eds.) 2005, *Proceedings from the Economic Evaluation of Occupational Health and Safety Interventions at the Company Level Conference*. Special Issue Journal of Safety Research - (36).

Ergoweb *web page:* http://www.ergoweb.com/resources/reference/history.cfm.

Forastieri V 1999, *Improvement of Working Conditions and Environment in the Informal Sector through Safety and Health Measures*. International Labour Office, Geneva.

Jastrzebowski W 1857, *An Outline of Ergonomics or the Science of Work Based upon the Truths Drawn from the Science of Nature*. Central Institute for Labour Protection, Warsaw, Poland.

Kirwan B 1994, *A Guide to Practical Human Reliability Assessment*. Taylor & Francis, London.

Koningsveld EAP, Zwinkels W, Mossink JCM, Thie X, Abspoel M 2003, *The Dutch national costs of poor working conditions* (in Dutch: Maatschappelijke kosten van arbeidsom-standig- heden van werknemers in 2001). Werkdocument 203, Ministry of Social Affairs and Employment, The Hague , The Netherlands.

Koningsveld EAP 2005, *Participation for understanding: an interactive method*. Journal of Safety Research - ECON Proceedings 36:231-236.

Koningsveld EAP 2006, Selling ergonomics: tricks or competencies? In: *Proceedings of the 13th Conference of the New Zealand Ergonomics Society*, Christchurch, New Zealand:1-5.

Marlow P 2006, *Making ergonomics a usual business tool; marketing ergonomic inter-ventions*. In: Proceedings of the Annual Conference of the New Zealand Ergonomics Society, Christchurch, New Zealand:39-42.

Oxenburgh M, Marlow P, Oxenburgh A 2004, *Increasing productivity and profit through health & safety*. Taylor & Francis, London.

Ramazzini B 1940, *From the Latin text of 1713, revised, with translation and notes by Wilmer Cave Wright*. University of Chicago Press, Chicago, USA.

Shapiro EC 1996, *Fad Surfing in the boardroom: managing in the age of instant answers.* Capstone, Knoxville, USA.

STAKEHOLDER DEFINITION AND CONTEXTUALISATION OF USABILITY CRITERIA IN HUMAN CENTERED PRODUCT DESIGN

Thomas Stüdeli [1,2]

[1]*Ergonomics and Product Design, Niederfeldstrasse 27, CH-8932 Mettmenstetten, Switzerland*
[2]*Biotronik AG, Center for Vascular Intervention, Ackerstrasse 6, CH-8180 Bülach, Switzerland*
E-mail: stuedelithomas@gmx.net, Phone: +41 76 331 36 20

Extended Abstract

Characterization of the user, specification of his/her needs and the actual use information are central themes in User Centered Design (UCD). In academia and practice there is an ongoing discussion on how to acquire the necessary user and use information and how and to what extend the user should be linked to the design and evaluation processes. Lately even well-established methods such as the usability evaluation have been questioned. Scott (2009, p.6) e.g. declared “actual usability methods can not keep up with computing” and proposes points for improvements for usability evaluations (2009, p.11), among others: “move away from lab tests to usability tests at the users site and in their context”, “making methods more contextual, fitting to smaller time frames and being able to deal with looser requirements”, and “do more collaborative research and participatory design, embracing the user as a designer”.

This presentation outlines an UCD methodology for the development of Medical Mixed Reality Systems (MMRS) (Stüdeli 2007). These complex medical devices support the interventional radiologist or surgeon during the intervention or operation with patient data and - in the case of robotic application - also with support of manual actions. The work environment in which these supportive medical devices are used is demanding and safety critical, work tasks are complex and often time critical. Two lessons learned from a four years development process in the light of the actual ergonomics standards are presented:

- The benefit of a stakeholder approach in comparison to a classical user approach. The stakeholder approach (Vink et al 2008) prevented us from focusing on the main users (intervention radiologist or surgeon). In an early stakeholder analyses the different role within the team have been described und used to select the right partners for our participatory approach. In a later stage the analysis turned out to be a good starting point to design systems functionalities for collaborative use.

- The benefit of contextualisation of usability and other evaluation criteria. Users and other stakeholders have been involved in the translation of typical ergonomic criteria such as efficiency, safety, workload, and comfort into the work context. For this contextualisation different stakeholders have been involved in different phases of the development. It was an additional way to transfer expert knowledge such as surgical experiences and know-how into the design process. Contrary to a common concern, this direct user involvement was very efficient and led to fewer and more precise criteria.

Documented usability engineering processes (IEC 62366:2007) and risk management processes (IEC 60601-1-6:2006) and the related involvement of medical experts are

mandatory to get market approvals for medical devices. Our experiences show that without contextualisation and stakeholder involvement the potential of these standards in the development process of a medical device is unfortunately not fully tapped. The industry can certainly benefit from incorporating a systems ergonomics point of view and a participatory ergonomics approach into these development processes. Blindly following a standard usability process might lead to inefficient amount of evaluation criteria and not lead to the expected benefits.

Our experiences endorse the introductory mentioned “points for improvement” from Scott also for demanding and safety critical work environments and complex tasks. Stakeholder analysis and can lead to fruitful participatory approaches “in the field” and introducing more contextual information can lead to more efficient usability evaluations.

Acknowledgements

During part of the described development process, the author has been Marie Curie Fellow under Marie Curie Actions for Human Resources and Mobility, contract MRTN-CT-2004-512400.

References

IEC 60601-1-6:2006: Medical electrical equipment - Part 1-6: General requirements for basic safety and essential performance - Collateral standard: Usability, International Electrotechnical Commission, Geneva, Switzerland.

IEC 62366: 2007: Medical devices – Application of usability engineering to medical devices, International Electrotechnical Commission, Geneva, Switzerland.

Scott, K.M. 2009. Is Usability Obsolete? Interactions. 16 (3): 2009. 6-11. http://dx.doi.org/10.1145/1516016.1516018

Stüdeli, T., Freudenthal, A. & de Ridder, H. 2007. Evaluation Framework of Ergonomic Requirements for Iterative Design Development of Computer Systems and their User Interfaces for Minimal Invasive Therapy, In: Allan Toomingas, Ann Lantz and Tomas Berns (Eds.) Proceedings WWCS 2007 Computing Systems for Human Benefits from the 8th International Conference on Work With Computing Systems. May 21-24 2007, Stockholm, Sweden.

Vink, P., Imada, A.S., & Zink K.J. 2008. Defining stakeholder involvement in participatory design processes, Applied Ergonomics. 39 (4): 2008. 519-526.

AGE AND GENDER RELATED MSD PREVELANCE DIFFERENCES IN SEDENTRY WORK

John Collins and Leonard O'Sullivan

Ergonomics Research Group
Centre for Physical Activity and Health Research
University of Limerick
Limerick

Abstract

As the world is entering an era of unprecedented demographic change, an inevitable outcome of our ageing society is the substantial increase in the number of people working later in life that has been the case up to now. Musculo- Skeletal Disorders (MSDs) are the most common work related health problem in Europe, and are likely to increase as exposure to work related risk factors for these conditions are increasing within the EU. This issue may be compounded by gender disposition where females are reported to be more susceptible to MSDs. Therefore, the purpose of this study was to investigate MSD prevalence and psychosocial risk exposures across working the age, separately for males and females performing sedentary computer work. The main finding of the survey was that prevalence generally increased with age, pointing to an "injured survivor" effect. Data did not indicate that the older workers and females in this study were exposed to more negative work-related psychosocial risks than were other workers.

Introduction

In 2008, The European Agency for Safety and Health at Work (OSHA) reported that Musculo-Skeletal Disorders (MSDs) were the most common work related health problem in Europe, affecting millions of workers. They stated that the size of the problem is likely to increase as exposure to work related risk factors for these conditions are increasing within the European Union (OSHA, 2008). Estimates of the cost of these problems are confined, however where data does exist the cost has been estimated at between 0.5% and 2% of Gross National Product (GNP) (Buckle and Devereux, 1999). Looking at it over a one year period in 2000 OSHA reported an estimated 350 million working days were lost due to work related diseases in Europe (OSHA, 2004). The HSE (UK) published results for a 12 month period between 2004/2005 and declared that an estimated 28.4 million working days were lost due to work related neck, shoulder, and back disorders. On average, each person suffering from these disorders took an estimated 23 days off work in that 12 month period, with an estimated cost to the economy of £5.7 billion. Averaged across the working population this represents an annual loss of 1.2 days per worker (HSE, 2006). The spiralling costs associated with absenteeism from work, and health care costs associated with MSDs, has fuelled interest in MSD research. There is a particular focus at present to maintain people in work throughout their working lives, and in particular to retain the older more experienced employees. This recent emphasis is in response to the ageing population worldwide. The European Union (EU) is facing unprecedented demographic changes that will have a major effect on various aspects of society. People are living longer and older people are enjoying better health and lifestyles.

By 2030, the number of older workers (55-64 years) will have risen by 24 million as the baby boomer generation become senior citizens. From 2005 to 2030 the number of people 65+ will rise by 52.3%, while the age group of 15-64 will decrease by 6.8% (EU, 2005). The ratio of dependent young and old people to people of working age will increase from 49% in 2005 to 66% in 2030; consequently to offset the loss of working age people there will need to be an employment rate of over 70%. The International Commission on Occupational Health (ICOH) acknowledged the increase of older workers and recognises that the survival of greater numbers of people into old age has great consequences for the health of the labour force, as many remain in employment with the deficiencies and disabilities that accompany ageing.

The EU's working age population is increasing and concerns are arising on how to control and keep older workers in employment. One of the main reasons that people leave the workforce is due to them experiencing work-related ill health (MSDs) or due to stress and depressive issues. The Commissions Green Paper, Faced with demographic change, a new solidarity between the generations, adopted on March 16th 2005 shows that from then until 2030 the EU will lack 20.8 million (6.8 pre cent) people of working age. Where the EU's working age ranges for 15-64 years. The ageing population gives rise to a particular challenge in ergonomics since the prevalence of MSDs increases with age; by their mid 30s most people have experienced their first episode of work related MSD, usually in the form of back pain (Bernard, 1997). As a person ages the resilience to MSDs decreases, with the loss of tissue strength leading to higher severity and a more frequent onset of soft tissue damage. Due to reduced resilience it has been shown in several studies that age is an important factor associated with MSDs (Guo et al., 1995, Soares et al., 2003, Parsons et al., 2007). Musculoskeletal impairments are among the most prevalent and symptomatic health problems of middle and old age (Buckwalter et al., 1993). In addition, many studies have indicated that females are more predisposed than males, especially in the upper limbs (Cook et al., 2000).

Age and gender-related risk factors range widely from anthropometric, household, hormonal and physiological variations. These risk factors may seem multifaceted; however, the complexity is reduced when the risks are characterised into biological and societal attributes. These phrases may be used in general terms to symbolise the varied number of risk factors into understandable titles. Nevertheless there is still much uncertainty on the true age and gender-related susceptibility to MSDs. Whether specific physiological or individual characteristics are more central over others are of yet unknown due to inconsistent findings, suggesting need for further constructive research concerning age and gender related risk. Uncertainty still remains on the susceptibility of the oldest workers to MSD development due to the phenomenon termed the Healthy Worker Effect. Other age-related physical reductions in strength, aerobic power, and work capacity may further increase older workers susceptibility to MSDs. The role that gender factors play in relation to MSD development identified increased prevalence to the female gender. Further, vulnerability may be due to the individual differences in physical capacities and coping strategies. Differences in skeletal muscle function related to ageing appear to be quantitative (muscle cell size and/or number), while for gender it appears to be qualitative (changes in single muscle fiber contractile velocity). Older individuals have an increased dominance of type I fiber types over their younger counterparts. Although it appears that male and females have similar overall distributions of type II and type I fibers, significant gender differences exist with regard to the total area occupied by each fiber type within muscle. It has been suggested that gender is of greater importance when considering anatomy and physiology parameters in relation to muscle fatigue.

The European Foundation for the Improvement of Living and Working Conditions have suggested that an emerging risk that will have an adverse affect on workers health, now and in the future, is exposure to increased use of computers. Generally, office based occupations are synonymous with low force static contractions, where current figures suggest that computers are becoming an increasing commodity in numerous organisations (Foundation, 2007). Due to the obvious and intense increase in computer use it is crucial to understand the effect that these devices have on employees' health and the direct effect low force static contractions have on MSD development.

The purpose of this study was to survey MSD prevalence and psychosocial risk exposures for older workers compared to younger workers, separately for males and females performing sedentary computer work, work which typically has high rates of shoulder/neck MSDs.

Method

Study purpose and design

The purpose of this study was to assess the level of ergonomic risk exposure in office based sedentary occupations and the associations with MSDs. The investigations were based on questionnaire data collected through an online survey in 2010. The survey was designed to identify age and gender-related associations to MSD development. The questionnaire surveyed various physical and psychosomatic aspects of work, including the psychosocial risk factors in sedentary occupations, and MSD development within this cohort. Furthermore, the exposure of office based employees to computer use and the level of occupational, physical and leisure activity were also captured.

Participants

The questionnaire was sent to the staff mailing list of two third level education institutions in the Republic of Ireland. In total 852 employees completed the online questionnaire. A prerequisite of participation in the survey was that each participant was employed within the organisation for 12 months or greater. This reduced the sample size to 709. All participants that displayed this condition were then considered for further analysis. The mean age was 40.25±11.15 years, mean stature 169.19±13.72cm, and mean body mass 71.26±15.28kg.

Questionnaire

The survey comprised three sections. Section (A) of the questionnaire was based on the Nordic Musculoskeletal Questionnaire (Kuorinka et al., 1987), to survey the prevalence of MSDs that participants experienced due to work. Section (B) surveyed physical activity levels based on the Baecke Questionnaire (Baecke et al., 1982). Baecke and associates developed the questionnaire to evaluate a person's physical activity, separating it into three distinct dimensions; physical activity at work, sport during leisure time, and physical activity during leisure time, excluding sport. Three corresponding indices arise from the results: the Work Index, Sport Index, and Leisure Index. The lowest possible value of the indices is 1.0 unit, signifying the lowest physical activity, while the highest possible value was 5.0units, signifying the highest physical activity. The Baecke Questionnaire is a simple, short questionnaire that is easy to self-administer making it a very attractive assessment tool for swift compilation of information. section (C) consisted of questions s1urveying psychosocial aspects at work, based on the short version of the Copenhagen Psychosocial Questionnaire (COPSOQ) (Kristensen et al., 2005). All three parts of the questionnaire were programmed into an online questionnaire via Survey Monkey ™ (Funchal, Portugal),

Results

Participant Details
The mean age for females was 40.31±10.89 years, mean stature 165.11±11.25 cm and mean body weight 65.85±12.37 kg. The mean age for men was 40.12±11.69 years, mean stature 177.86±14.47cm and mean body weight 82.74±18.53 kg.

Musculoskeletal Disorders
As to control for exposure to certain occupational risks the survey cohort were filtered. This entailed that individuals with similar exposure to PC usage were only considered for statistic analysis on MSD prevalence. Therefore this final cohort incorporated individuals who spent fifty percent or more of their workday in their office, and of this time spent in their office at least fifty percent was allocated to computer work (n=569). Figure 1 shows that neck (58%), shoulder (57%), and lower back (51%) are at highest susceptibility of injury. Figure 1 further affirms these regions as to high risk of injuries no matter the gender or age of the worker. Females consistently showed higher prevalence to all disorders, and in particular shows highest prevalence to neck (62%) and shoulder (62%) disorders. Figure 1 also identifies that males have greatest problems with their lower back (46%) followed by neck (41%) and shoulder (36%) region.

Taking the problematic regions separately, i.e. neck, shoulder, and lower back, Figure 2 identifies the age differences in susceptibility. All regions show minor differences, with neck disorders showing the only obvious increase across age (55%-58%-62%). Notwithstanding these minor fluctuations significant age differences were not apparent.

The problematic regions of the neck, shoulder and lower back were further analysed to identify the age differences within gender in more detail. Figure 3 shows the variation in prevalence across age, for both males and females. A negligible increase is apparent for neck (41%-43%-44%) and lower back (44%-49%-49%) among the males, while females show a decreasing trend for lower back disorders (59%-50%-50%). Independent sample t-tests were performed on males and females of similar age to identify if any gender differences were apparent for neck, shoulder, and lower back disorders. Significantly different gender differences were identified for each age group. Young (18-30) males and females reported significant differences in neck ($p<0.05$) and shoulder ($p<0.005$) disorders, while similar differences were reported for 30-50 year olds and the older cohort (51+) (Figure 3).

The young and oldest cohorts were compared as to identify differences in MSD prevalence. Figure 4 shows that the youngest and oldest groups report very similar prevalence, across both genders. For males, lower back disorders were at 44% prevalence for the 18-30 year olds and increased slightly to 49% for the oldest cohort (51+). For females, neck disorders increased from 64% to 70% for the youngest group to the oldest group.

Differences in risk exposure was investigated among all participants employed for the previous year or more (n=709). Table 1 show the gender and age differences among the cohort. Various significant age and gender differences are apparent, including highly significant age differences in PC usage at work ($p<0.0005$) and statistically high gender differences in office work ($p<0.0005$).

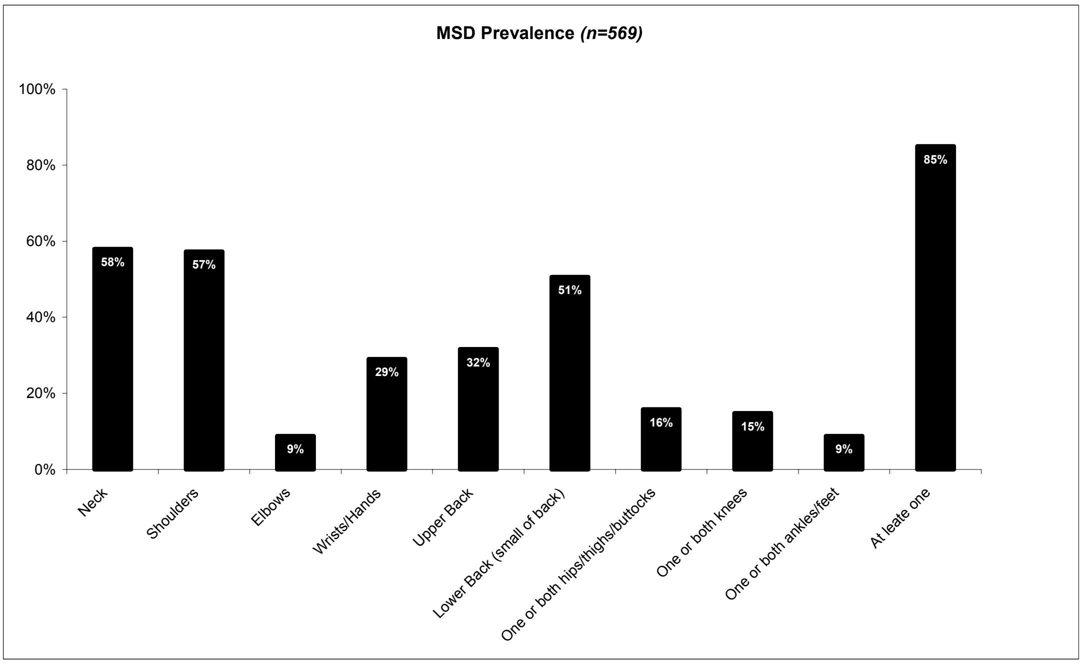


Figure 1 MSD Prevalence when controlling for PC Exposure (n=569)

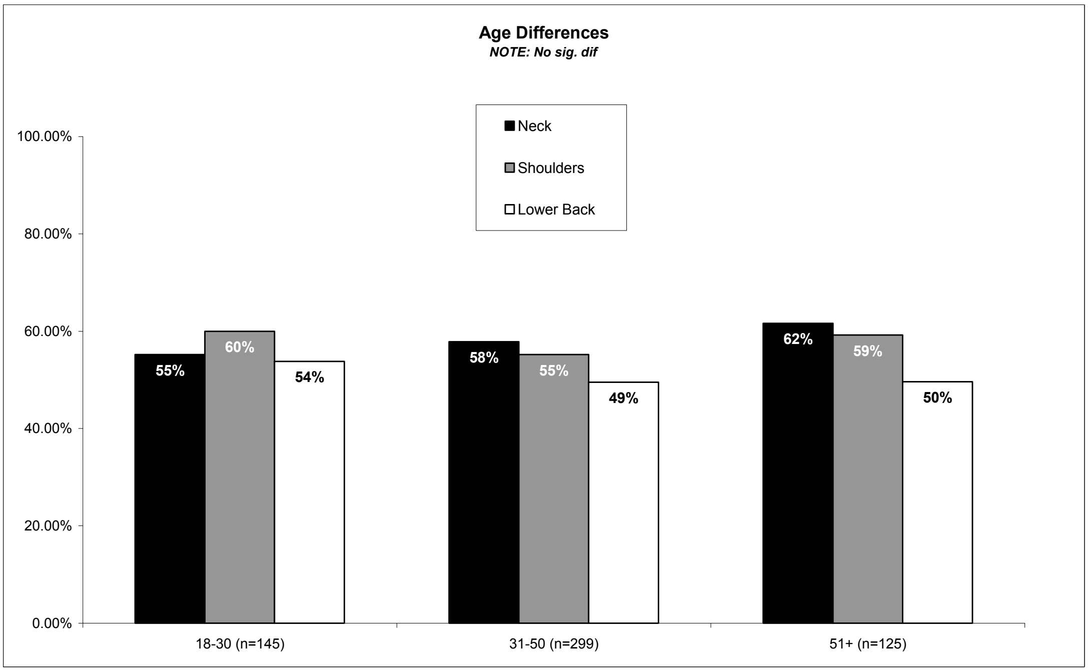


Figure 2 MSD Prevalence by Age (n=569)

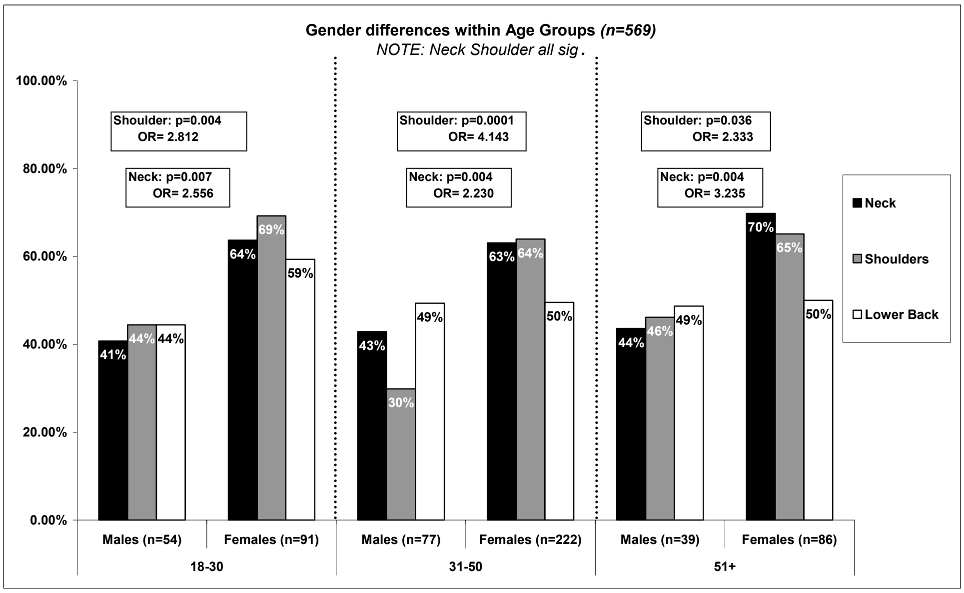


Figure 3 MSD Prevalence by Gender within Age Group (n=569)

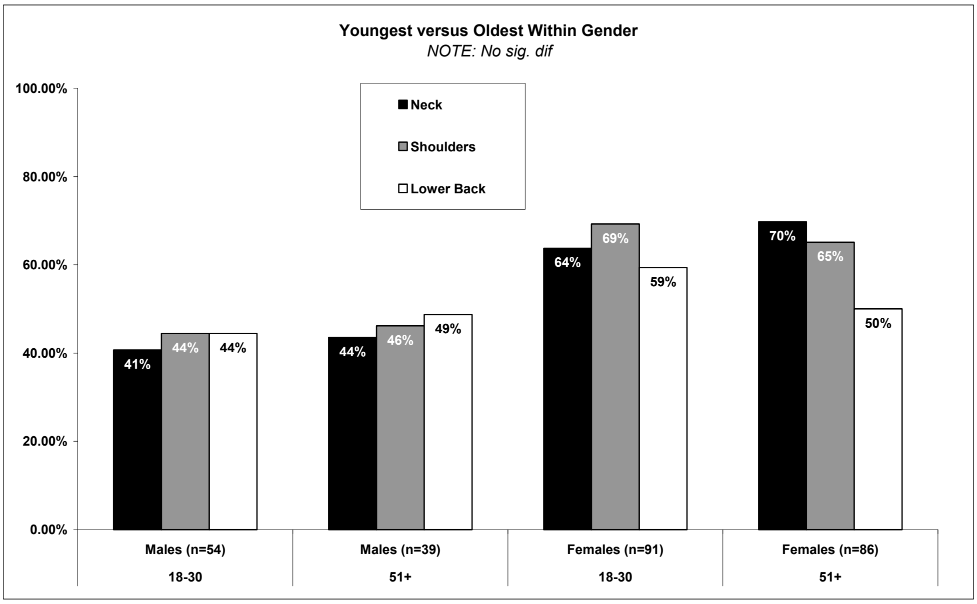


Figure 4 MSD Prevalence by Age Group within Gender (n=569)

Discussion

The biomechanical risk factors associated with computer work have been well studied, generally comprising prolonged periods of sitting with viewing of the computer monitor, and with sustained static muscle activity in the neck, shoulder and lower back regions. In addition, repetitive movements of the fingers and wrists in operating a keyboard or mouse pose a significant risk for cumulative trauma disorders. In 2007 Griffiths et al., identified that many professional workers are often encouraged to perform long hours of computer work with high mental demands, work at a hectic work pace resulting in heightened muscle tension and forces, and with inadequate work breaks. They further identified that as computerisation is intensified in workplaces it is important for organisations to identify and measure the risks to health and wellbeing associated with these changes.

Table 1 Age and Gender Related Differences in Exposure, Work/Sport/Leisure Index, and Psychosocial Risks *(n=709)*

Scale	Sig. [a]	Sig. [b]	Age	Male Avg.	Male SD	Male Sig. [b]	Male Sig. [c]	Female Avg.	Female SD	Female Sig. [b]	Female Sig. [c]
Employment (years)	p=0.744	p<0.0005	18-30	3.26	1.58	p<0.0005	$2^{\P}3^{\P}$	3.49	2.08	p<0.0005	$2^{*}3^{\P}$
			31-50	8.51	6.48		$1^{\P}3^{\P}$	8.83	6.46		$1^{\P}3^{\P}$
			51+	19.89	11.12		$1^{\P}2^{\P}$	16.13	9.84		$1^{\P}2^{\P}$
Office Work (%)	p<0.0005	p=0.698	18-30	0.75	0.22	p<0.0005	$2^{\P}3^{\P}$	0.69	0.23	p<0.005	$2^{*}3^{\P}$
			31-50	0.61	0.23		$1^{\P}$	0.74	0.24		1^{*}
			51+	0.60	0.25		$1^{\P}$	0.77	0.26		$1^{\P}$
PC Usage (%)	p<0.005	p<0.0005	18-30	0.80	0.21	p<0.0005	2^{***} $3^{\P}$	0.80	0.19	p<0.05	3^{*}
			31-50	0.68	0.26		1^{***}	0.77	0.19		
			51+	0.62	0.26		$1^{\P}$	0.71	0.26		1^{*}
Work Index	p<0.05	p=0.647	18-30	2.40	0.41	p=0.239		2.43	0.31	p=0.321	
			31-50	2.49	0.47			2.35	0.44		
			51+	2.41	0.44			2.32	0.51		
Sports Index	p<0.0005	p<0.0005	18-30	3.23	1.03	p<0.0005	$2^{*}3^{\P}$	2.74	0.88	p<0.0005	$2^{**}3^{\P}$
			31-50	2.79	1.00		$1^{*}3^{*}$	2.45	0.90		$1^{**}3^{*}$
			51+	2.45	0.86		$1^{\P}2^{*}$	2.24	0.93		$1^{\P}2^{*}$
Leisure Index	*0.413*	p=0.491	18-30	2.63	0.74	p=0.722		2.64	0.60	p=0.533	
			31-50	2.65	0.64			2.72	0.65		
			51+	2.71	0.65			2.73	0.74		
Job Content	p<0.0005	p<0.05	18-30	71.23	12.68	p=0.733		65.40	15.66	p=0.063	2^{*}
			31-50	67.78	17.10			60.53	19.02		1^{*}
			51+	68.44	17.47			62.02	20.24		
Job Demands	*p=0.856*	p=0.496	18-30	71.23	12.68	p=0.888		54.84	15.14	p=0.398	
			31-50	55.30	16.93			53.56	18.74		
			51+	54.76	18.10			55.73	20.00		
Work Environment	*p=0.223*	p=0.242	18-30	56.35	14.95	p=0.505		59.44	17.50	p=0.177	
			31-50	55.38	16.05			54.58	18.89		
			51+	51.79	17.26			56.06	20.07		

** $p < 0.01$ (2-tailed)
* $p < 0.05$ (2-tailed)
¶ $p < 0.0005$ (2-tailed)
*** $p < 0.005$ (2-tailed)

1 = Sig. with 18-30
2 = Sig. with 31-50
2 = Sig. with 51+

a. Gender difference in exposure
b. Age difference in exposure
c. Age comparisons

Whether working at home or in the workplace, using computers is an increasingly common part of employment. These jobs can be relatively varied and interesting but may also involve repetitive tasks and long periods in the same physical position. In the current study males and females did not only differ in exposure to office based work, but additional significant gender differences were identified in relation to computer use. These gender disparities, in this case within two white-collar organisations, show similar trends with many of our European neighbours. The Fourth European Working Conditions Survey (Parent-Thirion et al., 2007) identified significant gender differences in computer use across occupations, with white-collar work having the highest rate of regular computer use for both women and men. For women, managerial jobs have the highest rates of computer use, while for men the highest rates occur in professional occupations. A somewhat similar picture emerges for use of the internet at work although the gender gap is smaller. Unsurprisingly, white-collar occupations are those most likely to use the computer regularly, that is, more than 50% of the time. Again in this study women were more likely to be regular users overall, however international statistics show that male professionals and blue-collar operators and labourers have higher usage than their female counterparts. Thus, although women are more likely to make use of computers and the internet, the occupational and working time distribution of women's and men's jobs may play a significant role in this gender disparity (Burchell et al., 2007).

In relation to computer use by age, significant age differences were identified across the entire cohort and within gender. Generally computer exposure declined with advancing age, particularly for males. Female prevalence levels remained high throughout the age groups suggesting the higher requirement of computer use within this cohort. According to the European Working Conditions Survey (EWCS) (2008), in all age groups, women reported greater use of computers at work than men.

Psychosocial risks did not show strong associations to age and gender differences in MSD development in office based occupations. The lack of significantly strong associations strengthens the belief that psychosocial risks are weak in identifying or exacerbating MSD development. Nevertheless, stress in its original form may have a high association with occupational disorders; even if psychosocial risks are poor in encapsulating these associations

Conclusions

Musculoskeletal disorders in the form of lower back, neck and shoulder disorders were common among these sedentary workers. Disorders generally increased with age which adds further strength to an "injured survivor" effect, yet age by itself was not associated with MSD development. Gender was the strongest predictor of neck/shoulder disorders with females reporting elevated risk in each of the age groups. The broad term of 'computer exposure' did not exclusively explain the increased susceptibility to neck, shoulder and lower back disorders, however the more specific biomechanical and psychological inadequacies associated with such work, namely excessive keyboard use and mental tiredness, may identify significantly stronger associations with musculoskeletal symptoms. However further research is needed to substantiate this suggestion.

Psychosocial risks did not show strong associations with MSD reports in this office based occupational group.

Lack of activity at work is a major concern in relation to MSD susceptibility as it was lower for females and reduced with age, and negatively correlated with MSDs. The positive effect on maintaining a dynamic and active existence during work and leisure suggests the introduction of activity regimes to keep our ageing population occupationally active and free from work-related disorders. However, this might be considered an occupational health issue rather than an ergonomic issue.

References

BERNARD, B. P. 1997. Musculoskeletal disorders and workplace factors: A critical review of epidemiological evidence for work related musculoskeletal disorders of the neck, upper extremity, and low back. In: NIOSH (ed.). U.S. Department of Health and Human Services, Cincinnati, USA: NIOSH.

BUCKLE, P. W. & JASON DEVEREUX, J. 2002. The nature of work-related neck and upper limb musculoskeletal disorders. Applied Ergonomics, 33, 207-217.

BURCHELL, B., FAGAN, C., O'BRIEN, C. & SMITH, M. 2007. Working conditions in the European Union: The gender perspective. Dublin: European Foundation for the Improvement of Living and Working Conditions.

COOK, C., BURGESS-LIMERICK, R. & CHANG, S. 2000. The prevalence of neck and upper extremity musculoskeletal symptoms in computer mouse users. International Journal of Industrial Ergonomics, 26, 347-356.

EUROSTAT. 2008. Annual review of working conditions in the EU 2006–2007 http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1090,30070682,1090_33076576&_dad=portal&_schema=PORTAL

FOUNDATION 2007. Fifteen years of working conditions in the EU: charting the trends. Dublin: European Foundation for the Improvement of Living and Working Conditions.

GRIFFITHS, K., MACKEY, M. & ADAMSON, B. 2007. The Impact of a Computerized Work Environment on Professional Occupational Groups and Behavioural and Physiological Risk Factors for Musculoskeletal Symptoms: A Literature Review. Journal of Occupational Rehabilitation, 17, 743-765.

GUO, H. R., TANAKA, S., CAMERON, L. L., SELIGMAN, P. J., BEHRENS, V. J.,GER, J., WILD, D. K. & PUTZ-ANDERSON, V. 1995. Back pain among workers in the United States: national estimates and workers at high risk. American journal of industrial medicine, 28, 591-602.

HSE 2006. Health and Safety Statistics 2005/2006. HSC. EU 2005a. Green Paper - Confronting demographic change: a new solidarity between the generations. Brussels: Commission of the European Communities.

KRISTENSEN, T. S., HANNERZ, H., HOGH, A. & BORG, V. 2005. The Copenhagen Psychosocial Questionnaire - a Tool for the Assessment and Improvement of the Psychosocial Work Environment. Scandinavian Journal of Work, Environment & Health, 31, 438 - 449.

OSHA 2008. Prevention of work-related musculoskeletal disorders in practice. Luxenbourg: European Agency for Safety and Health at Work.

PARENT-THIRION, A., FERNÁNDEZ MACÍAS, E., HURLEY, J. & VERMEYLEN, G. 2007. Fourth European Working Conditions Survey. Dublin: European Foundation for the Improvement of Living and Working Conditions.

PARSONS, S., BREEN, A., FOSTER, N. E., LETLEY, L., PINCUS, T., VOGEL, S. & UNDERWOOD, M. 2007. Prevalence and comparative troublesomeness by age of musculoskeletal pain in different body locations. Family practice, 24, 308-316.

SOARES, J. J. F., SUNDIN, O. & GROSSI, G. 2003. Age and musculoskeletal pain. International journal of behavioral medicine, 10, 181-190.

THE USE OF AUGMENTED REALITY TO SUPPORT MAINTENANCE: HUMAN FACTORS ISSUES AND ADVANTAGES

M.C. Leva[1], A. M. Kay[1], T. Smith[2], A. Diez[3], J. Azpiazu[3] , P. Liston[1], S. Cromie[1]

[1]*School of Psychology University of Dublin Trinity College Dublin Ireland.*
[2]*NEM Solutions UK The TechnoCentre, Puma Way Coventry, UK*
[3]*Tecnalia, Paseo Mikeletegi 7, Parque Tecnológico, 20009 Donostia-San Sebastián, España*

Abstract

The ManuVAR project, funded under the European Union 7th Framework Programme, is a response to changes in the European manual-labour market. In the past two years the project has been attempting to exploit the potential of virtual reality (VR) and augmented reality (AR) to improve communication between people and systems to support train maintenance operations. The purpose of the application being explored is to facilitate technical trouble shooting, more effective delivery of instructions across different locations for companies competing on international markets with limited personnel resources and availability to travel, and the reuse of the solutions identified for recurring similar interventions. This paper aims to describe how the solutions identified within ManuVAR can achieve these goals and the HF challenges being addressed already within the project to support the initial developmental phase.

1 ManuVAR and the need for support of manual work in Train Maintenance

The ManuVAR project, funded under the European Union 7th Framework Programme, is a response to changes in the European manual-labour market (ManuVAR 2011). High-value, high-knowledge manual work is essential to European competitiveness – Train vehicle maintenance is a critical area in this respect and there is a necessity to keep high standards and still drive costs down in order to keep a competitive profile in the market considering the role played by developing-nations where labour costs are lower but where reliability, quality and efficiency are more difficult to guarantee.

Within ManuVAR in the past two years there has been an attempt to exploit the potential of VR and AR to improve communication between people and systems to support train maintenance operations.

The purpose of the application being explored is to facilitate technical trouble shooting, more effective delivery of instructions across different locations for companies competing on international markets with limited personnel resources and availability to travel, and the reuse of the solutions identified for recurring similar interventions.

Train vehicles are complex and made up of hundreds of subsystems: electrical, pneumatic, and mechanical; and in many cases provided by different manufacturers and system suppliers. A remote maintenance set-up would enable a more seamless process to communicate with experts located in equipment supplier headquarters, possibly remote from the real operations. Furthermore the possibility to manage and handle the maintenance actions performed and the possible solutions identified is also an added value to the current situation, so that they can be stored and accessed for additional consultations by train designers and maintenance experts.

This paper aims to describe how the solutions identified within ManuVAR can achieve these goals and the HF challenges being addressed already within the project to support the initial developmental phase.

2 Manual work in Train Maintenance: the "status quo"

Thanks to recent advances in broadband networking "technology already evident in businesses, households, and cities, more and more systems are being interconnected and all sorts of new services are being made available at an accelerated pace. There is a strong expectation and desire that railroad systems can similarly be made more efficient and convenient by using these same technologies to interconnect many onboard train and ground systems" (Ishida et al 2004).

However the manual element in it is still predominant. Most of the vehicle maintenance interventions take place in depots managed by the rolling stock maintainer, only emergency corrective maintenance occurs in stations or on track (ManuVAR Deliverable 1 2009).

In the last twenty years the working conditions and installations in maintenance depots have improved dramatically, due to trends in privatization and the need to provide a service of highly reliable trains, at low operating cost, and with a high level of availability and safety.

Despite significant changes in working conditions, due to the nature of train maintenance, most corrective maintenance often takes place on equipment that is not directly accessible and is mounted in a confined space. Parts of the equipment are often hidden to view, and it is often difficult for the maintainer to have documents and drawings close at hand when attending the faulty equipment.

The Main Issues connected with the current situation.

Unplanned maintenance (and therefore unavailability of the train vehicle to operate) is due to unexpected corrective maintenance actions following accidents or unexpected findings during routine checks.

In the event that an accident or unexpected fault is the source of intervention (and from field data studies this often appears to be the case, see paragraph 4 for more details), the train is normally brought to the nearest depot, where it is not always possible to have the most suitable expert to solve the problem immediately at that site. In addition, the dependence on expertise from equipment suppliers, also limits the competencies of maintenance staff to immediately diagnose and fix a problem. Thus, the support of an expert either from the maintainer's headquarters or from an external provider is often required.

Firstly, engineers must search internally to see if their required competencies are available to get a diagnosis and a solution. If this is not the possible, then phone communication may be established with headquarters. An expert is identified, either within the organisation or with the equipment supplier, and is typically to the depot, or can be sent to the external location if the train is out on track. The expert must bring all the tools and spare parts that they believe may be required for the job.

Delays and associated costs in this process are evident. Experts may not be available immediately to travel to site, travel plans must be organised, and then when at the site, the expert may not have taken with them the correct tools or parts. In addition, he/she may not be

the appropriate expert if the fault is subsequently associated with different equipment. Maintenance histories of potentially affected subsystems and recent condition reports may be hard to obtain.
A visual mapping of the processes that may occur for a corrective maintenance intervention as the one outlined above is reported in Figure 1.

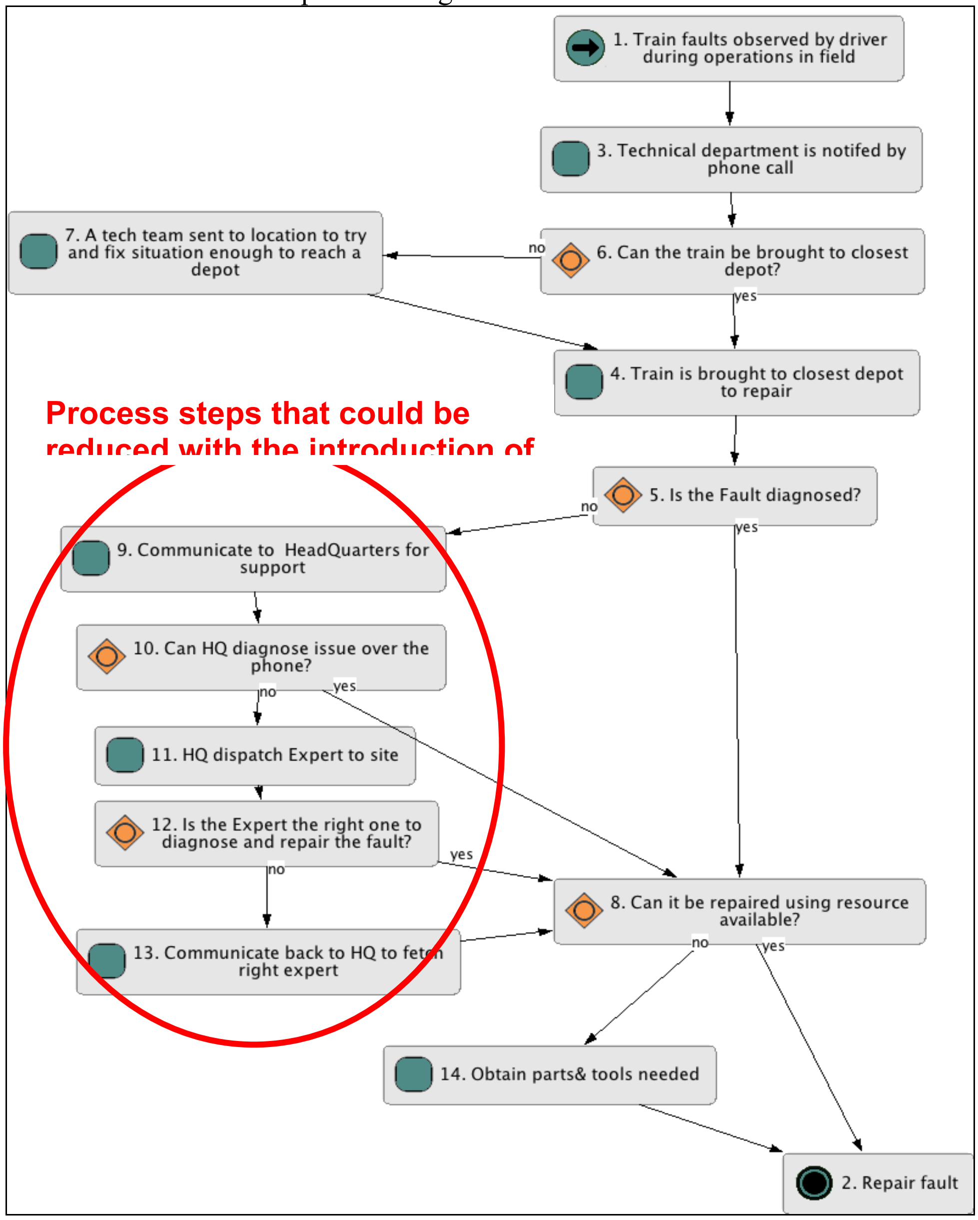


Figure 1: Map of the process followed for corrective maintenance in the current situation

3 The ManuVAR approach: challenges and opportunities.

How can Maintenance be supported Using AR: a brief description

AR techniques aim to convey information to the user which is spatially coherent with the observed scene. They display such information by augmenting the scene captured through a camera with graphical objects properly aligned with the real world 3D structures (De Crescenzio et al .2011).

In ManuVAR, the proposed AR based application: for enhanced maintenance support should be able to perform the following functions:

- The worker should be supplied with a computer. He/She can wear a Personal Mobile Device (PMD) i.e. the goggles, with a web camera attached to them and there is a tracking element and AR content rendering element running on this computer:
- The application tool should be able to identify visually (through the information captured by the web camera, an element, even if made up of several subcomponents, depending on what the worker is looking at and previously stored virtual objects (e.g. CAD-CATIA file)
- The metadata information associated with each subcomponent can be made available to the worker through the screen of the computer or the goggles and overlaid on the actual system using labels and text messages.
- The system should be able to establish and support a communication between the worker and a remote expert through the web. Relevant information to be provided (e.g. diagnostic information collected through the on board electronic system) related to the object and/or component identified by the system should be made available through a connected database.
- It should then provide the possibility to save the workflow information and tasks performed on the case study in a specified troubleshooting database for further consultation.

Thus, two main types of instruction can be given:
- Meta-information overlaid over worker's view and associated with the object and/or component identified by the system (CAD diagrams and models, diagnostic information and reports, instructions of the manufacturer manuals etc.)
- Voice instructions, provided by a remote expert

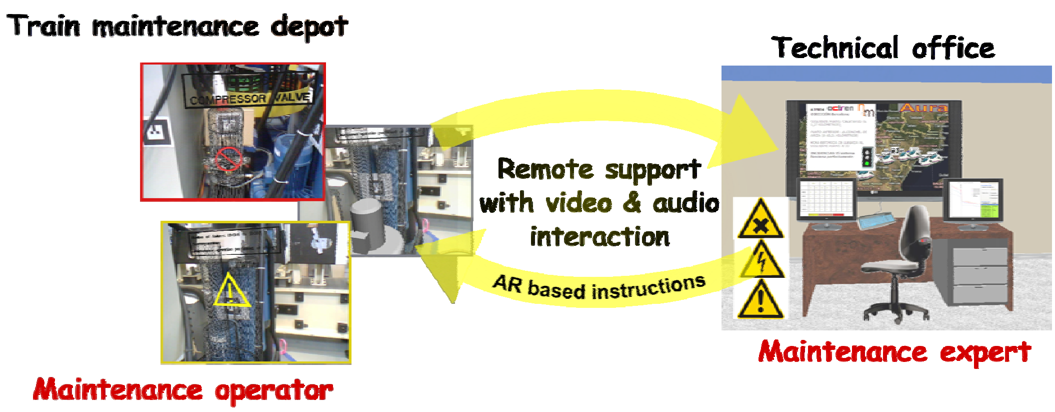


Figure 2: Schematic Representation of the application being developed within ManuVAR

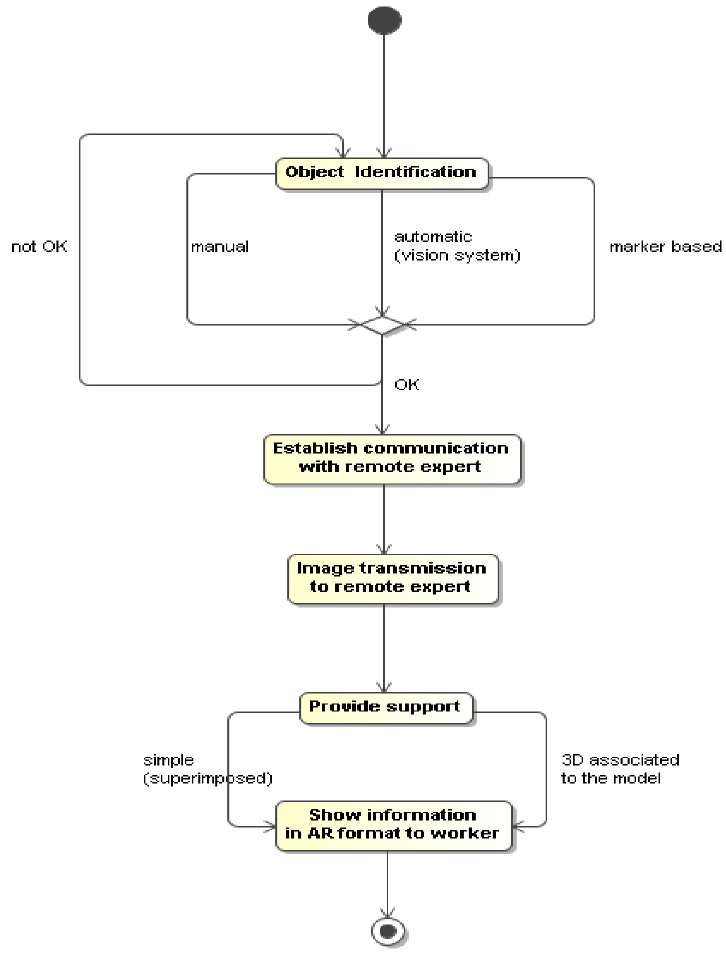


Figure 3: AR remote support/maintenance State Machine Diagram

Current Trials and validation of the application.
ManuVAR follows a user-centred participatory design approach (Vink et al. 2005) which actively involves the relevant stakeholders in the design and evaluation process. It is a way of developing a socially valid and responsive system, ensuring the technology matches the needs of the user, increasing agreement on the final solution and proving a greater sense of ownership.

Within ManuVAR, every developmental stage is marked by a trial aimed at evaluating the functionalities expected for the current developmental stage. The observation of how the product satisfies best practices for VR and AR usability conditions, so as to provide an interim feedback loop in the developing cycle is also evaluated.

Specifically during each trial, the interim development of the application was evaluated in three ways:

[1]. The application was tested against the subset of Functional Requirements previously specified with the end users and selected for the current developmental stage for the application at that time.
[2]. Users are tested with a short checklist for motion sickness and comfort before and after the trial
[3]. An interview and an observation protocol are also used to provide the end user and the Human factors expert feedback organized around the following sections:
    a. Setting up the task,
    b. Performing the task

c. Display of task progress
d. Accessing and storing data
e. Visualising the data
f. General Performance

In the sections presented above, the observers and the end user were able to provide an evaluation of the following non functional requirements issues:

a) General performance of the application
b) Navigability
c) Match between system and real world
d) User comfort
e) Visibility of System Status
f) Errors Correction
g) Workload
h) Communication

The HF challenges and potential improvements of using AR.

In relation to the current situation illustrated in Chapter 2, the users of the tool pointed out the following potential advantages:

1. The effective combination of data coming from the train's diagnostic system and an AR system could greatly improve maintenance tasks:
    a. An expert may guide the maintenance operator from a remote location, helping him to focus on the most important points of the equipment;
    b. The diagnostic system can suggest or refer to locations of where the anomaly may potentially be, as well as providing contextual information for on-site fault analysis;
2. AR systems can provide details such as 2D drawings or 3D models to assist the maintainer in understand the equipment being examined.
3. A AR/VR system can be used in conjunction with web-accessible maintenance information such as maintenance task records, fault histories and recent diagnostic reports.
4. A video appraisal of the situation could actually provide a better indication on what may be needed to repair a damage in case spare parts have to be ordered and an expert may still be needed on site. Thus shortening the possible delays due to not having the right equipment or parts on site.
5. There is an added potential of VR systems to provide a form of training to staff, in order to improve competence levels prior to undertaking work, thus reducing risk and improving time to repair.
6. The use of AR based tools to connect to the suppliers of supplier of very specific types of equipment (e.g. Air Conditioning systems etc.) can also be considered as a major improvement.

Augmented reality (AR) may well be a promising technology to build advanced interfaces using interactive and wearable visualization systems to implement new methods to display documentation as digital and graphical data. However, there are a number of issues to be addressed and from trials already performed within ManuVar. The following were identified:

- the possible use of cumbersome hardware (the goggles can be heavy and may need to be wired to the computer)
- the need to put markers on the parts being analysed to align the AR model to the real objects.
- the need to quickly create digital content to apply to the AR model or to the real one is a complex technological task

All the above issues seem to hinder its effective implementation in the industry as confirmed also by previous attempts (De Crescenzio et al 2011, Regenbrecht et al 2005). However the use of mark less camera pose (markerless tracking) estimation is becoming more and more of an attractive alternative solution.

In dealing with the difficulty of sending quick digital content (instructions, arrows etc.) a possible solution is the alternative use of “offline and an online process”. According to what described by De Crescenzio et al. (2011) in the offline process it is possible to acquire a reference image of the object that has to be augmented and extract local invariant features.
This information can then be stored and extracted features in the system can be used whenever needed (e.g. it is easier to daw and position a circle on a static image than on the video streaming of the real object itself). Figure 4 and Figure 5 shows two images taken form one of the trials performed using the application in March 2011.

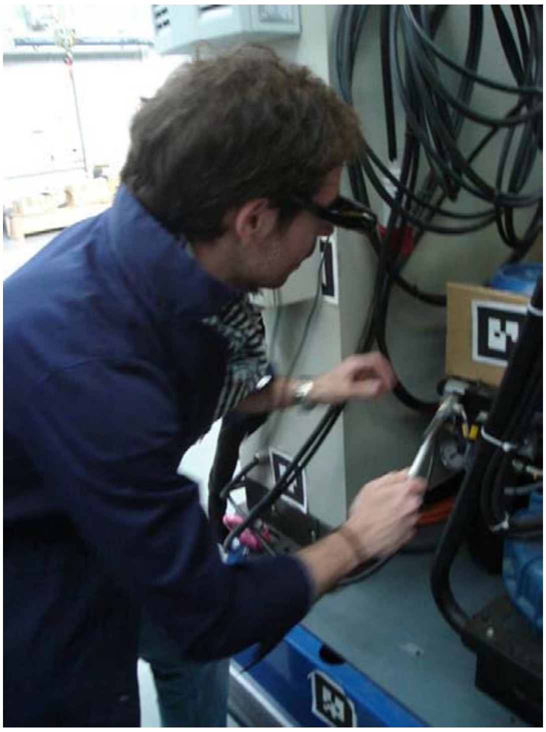


Figure 4: Worker observing the part with goggles, the screen on the left hand side displays what the worker sees through them.

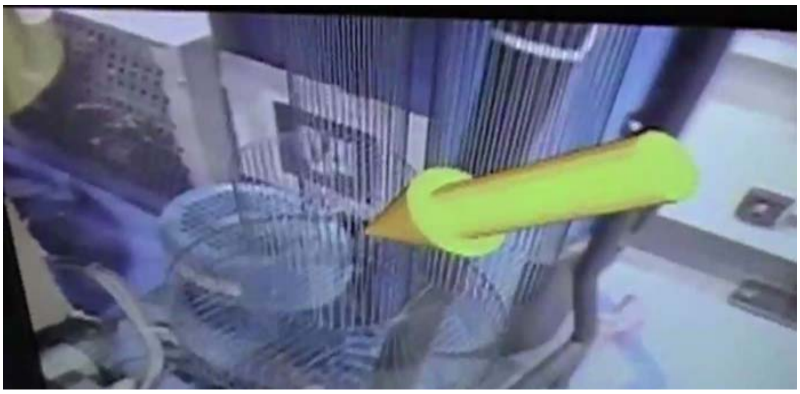


Figure 5: Closer view of what the worker sees through the goggles with instruction provided in AR.

4 Preliminary evaluation of the potential benefits of the envisaged solution

The possible future process for corrective maintenance using the support of the AR proposed solution can be represented in Figure 6. It is immediately evident by comparing this process to the one used for describing the current situation that the number of steps to go through in order to reach a solution become significantly less (especially in the most time consuming cases of unplanned interventions) if the remote support connection can be established. In this case in fact all the logistics costs and time consuming activities connected to the need of sending an expert on site to asses the situation are going to be dramatically reduced.

Table 1. Summary of business case for MX operator AR support (ManuVAR Del. 2 2009)

Business use case	AR remote support/maintenance
Business use case description	The description of this use case concerns the remote on-line based AR support provided by a maintenance expert located in another depot or in a technical office
Actors	• Maintenance expert • Worker
Performance goals	• Reduction of overall maintenance costs • Time and costs reduction derived from travelling expenses and logistics • Reduced need for invasive maintenance procedure • Reduction of less maintenance induced damage • Improve the availability and performance of trains • Reduction of spare parts consumption
Preconditions	A maintenance activity must be performed over a train component, AR system is installed and running and remote communication is available
Post conditions	The maintenance activity should be completed and recorded by the system for future usage
Scenario	Worker is in depot, carrying the AR equipment, and maintenance expert is in his/her workstation with the AR software and hardware installed and running: • Worker comes up to the component he/she is going to work with • The component/object is identified: - manually - marker based - automatically, by means of a vision system • The link with remote expert is established • captured real-time image by worker camera is transmitted to remote expert • Maintenance expert offers his/her support by voice chat • Expert provides the information that will be shown in AR format to worker: - AR information is superimposed to real image - 3D information is associated to object model

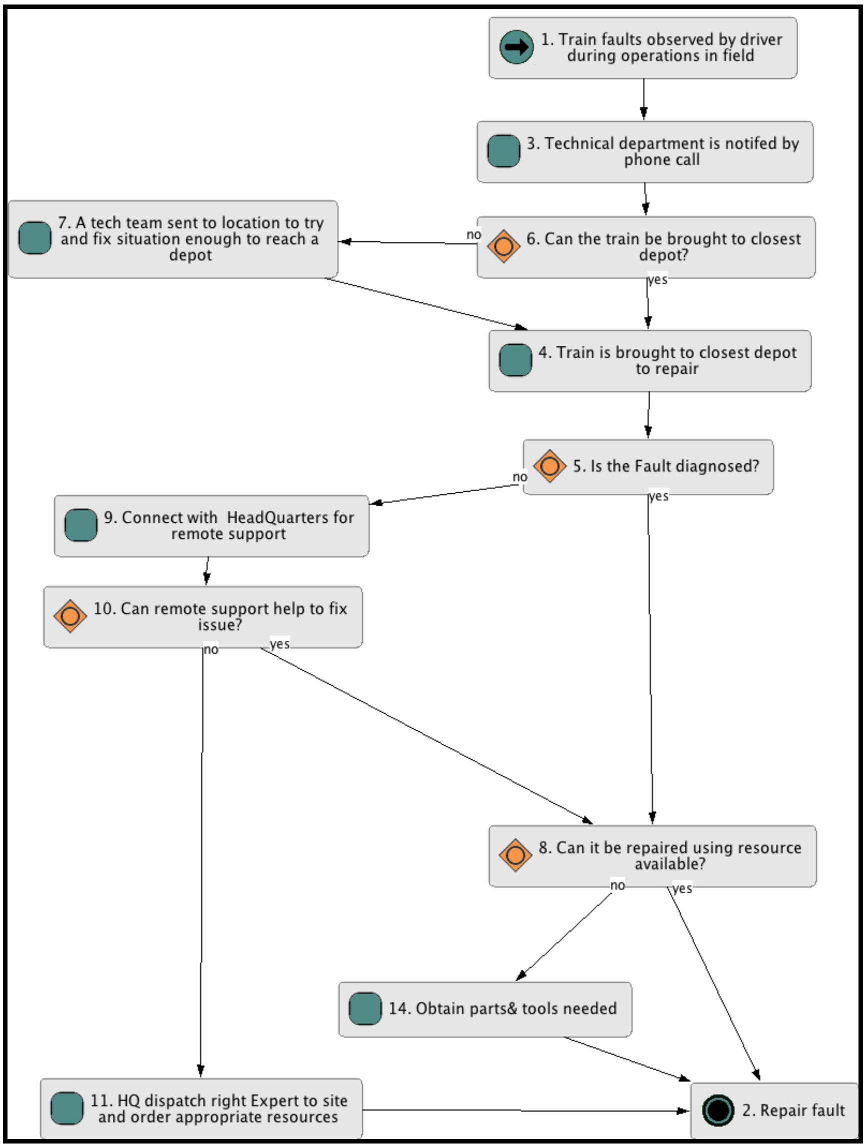


Figure 6: Map of the process that could be followed in the future remote support situation

In a study carried out in Estonia for Light Rail it was found that "during the last 14 years the direct impact of maintenance efficiency on the failure rates and repair costs can be very significant. The main causes for traffic stops were technical failures and accidents; 75% (53) of downtime was caused by them. 25% of downtime was caused by power department, trail service and other (mainly for reconstruction work)" (Rosin et al. 2006).

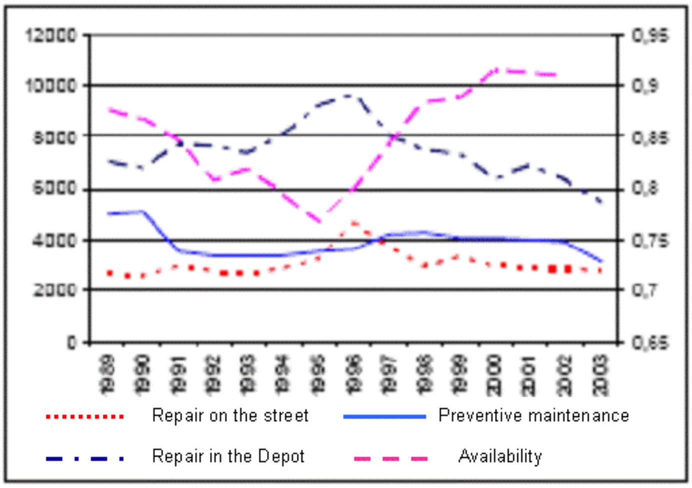


Figure 7: Repair, maintenance, availability as it emerged from the study of Rosin et al. (2006)

From the study it emerged that a predominant part of the service and maintenance work there are repairs on request - 42%, with repair on the tracks being also another 21% of the total, while preventive maintenance and reconstructions constitute only the 27% and 10% respectively of the total (Rosin et al 2006) (see Figure 7).

Shorter and more efficient corrective maintenance interventions therefore mean a significant increase of train availability. All these translate into a reduction in the maintenance costs and an improvement in reliability and passenger comfort.

5 Conclusions

The users of the application tested within the trials of ManuVAR gave very positive feedback about the application and would appreciate the possibility of implementing the prototype on a real scenario and the possibility of including it as part of the available troubleshooting methods and tools for corrective maintenance intervention. This could be as a result of the technology actually enhancing the task's efficiency.

This system can demonstrate AR's potential in assisting a more effective transfer of knowledge and information between a technician and an expert who are remotely connected. However, there are still some stumbling blocks to be effectively targeted for a much more user-friendly version of the technology, i.e. the use of markerless methods with an improved user-friendly graphics-authoring procedure. This would enable wider applicability and could be transferred to different pieces of equipment in a simpler and more effective manner.

Acknowledgements
The above mentioned research has received funding from the European Commission's Seventh Framework Programme FP7/2007-2013 under grant agreement 211548 “ManuVAR”.

Reference

De Crescenzio, F. ; Fantini, M. ; Persiani, F. ; Di Stefano, L. ; Azzari, P. ; Salti, S.(2011) Augmented Reality for Aircraft Maintenance Training and Operations Support Computer Graphics and Applications Volume:31 Issue:1

Ishida K., H. Kitabayashi, M. Nagasu, K. Bekki, “New train control and information services utilizing broadband networks,” Hitachi Review, Volume 53, No. 1, February 2004

ManuVAR (2009) Deliverable 1.1 v1 20-11-2009 “Definition of requirements, analysis of standards and tools” ManuVAR internal Document CP-IP-211548.

ManuVAR (2009) Deliverable 2 v1 9-12-2009 Definition of industrial cases ManuVAR internal Document CP-IP-211548.

ManuVAR (2011). ManuVAR website homepage. [online] at: <http://www.manuvar.eu>
[Accessed 28 of May 2011].

Regenbrecht H., G. Baratoff, and W. Wilke,(2005) Augmented Reality Projects in the Automotive and Aerospace Industry. IEEE Computer Graphics and Applications, vol. 25, no. 6, 2005, pp. 48–56.

Rosin A., Moller T., Laugis J.,(2006) Maintenance Analyses, Control and Supervision System Design of Light Rail – An Application in Estonia. Proceedings of the 32nd IEEE annual conference on Industrial Electronics, 2006, 489-494.

Vink, P., Nichols S., and Davies R.C. (2005). Participatory Ergonomics and Comfort. In P.
Vink (Eds.), Comfort and Design (pp. 41-54). Florida:

THE CHALLENGES OF ORGANIZATIONAL CHANGE: WHAT HAS HUMAN FACTORS TO OFFER

N. McDonald[1], M.C. Leva[1], S. Corrigan[1], P. Ulfvengren[2], D. Baranzini[1]

[1]*Aerospace Psychology Research Group School of Psychology University of Dublin Trinity College IE*
[2]*KTH Industriell teknik och management Address SE-100 44 STOCKHOM (Sweden)*

Abstract

Within the aviation industry the need for sustainable change is becoming more and more unavoidable. Change is in fact something being imposed on the industry from a number of sources (e.g. Upcoming Regulations, in-creasing commercial pressure and the introduction of new technologies) and the need for change management skills and capability within organisations is increasing. The paper presents the initial approach to answer those needs in a research project called MASCA (MAnaging System Change in Aviation. The project is sponsored by European Union to deliver a structure to manage the acquisition and retention of skills and knowledge, through training on organisational processes for managing change. MASCA takes an action research approach with a primary focus on the transfer of change management capability into the organisations that are responsible for and involved in change initiatives; especially in the area of performance management in safety.

1 Introduction: the need for Change in Aviation

The Advisory Council for Aeronautics Research in Europe (ACARE) vision for the transformation of the aviation system by the year 2020 demands that the system delivers to the customer – the travelling public and, more broadly, society – large step changes in quality, safety, security, cost and environmental impact (ACARE 2001). All of these are to be achieved while coping with greatly increased capacity. While new technologies are essential to achieving these objectives, these are all operational goals where the improved outcome for the community is delivered by a transformed operational system. For this reason, managing change has to be at the centre of achieving the realization of the ACARE vision.

Commercial competition, driven partly by the 'low-cost' business model, is driving aviation organizations globally to change the way in which they do business, cutting costs and developing a leaner enterprise. The low-cost carriers, relatively new entrants to the business, grew their companies around this business model. Older organizations (including the 'legacy carriers') do not have the opportunity to build their organization from scratch – therefore they have to change an established stable system with its strong cultural and institutional sup-ports, without compromising core operational goals.

At the same time, in the operational sector, new requirements for regulation coming from ICAO (2009) demand that all aviation organisations develop a safety management system, requiring transparency in the development of internal organisational processes. The requirements for safety management require a proactive strategic approach, anticipating risks and demonstrating a capacity to keep safety at the centre of change that is driven by commercial competition, and ensuring that safety evidence itself becomes an effective driver of change, even in the 'ultra-safe' system that aviation has become.

In MASCA the opportunities offered by Human Factors expertise is directed towards achieving the right acquisition and retention of skills and knowledge, through training on organisational processes for managing change. This in fact is on of the main needs highlighted from personnel involved in the current situation across the whole air transport system. The project includes different stakeholders in a common operational system (airlines, airports, maintenance companies, etc.) in the common effort to identify critical areas to change the shared operational system to deliver a better service, especially in the area of performance management in safety.

The present paper will discuss the main drivers for change in aviation, the challenges and issues that any change initiative in this context may need to face, and the main Human Factors principles upon which the project is attempting to achieve its objectives.

2 The high failure rate of organizational change

The literature on organizational change demonstrates that, against different criteria and outcomes, only a minority of major change initiatives (typically between 30% and 50%) have a positive outcome (Dent and Powley, 2001, Porras and Robertson, 1983, Kotter, 1995, Pascale, Millemann and Gioia, 1997, Maurer, 1996). From some of the few longitudinal studies of change, Pettigrew shows how change is complex, frequently opportunistic, and depends on the balance of capacity within the organization with the opportunity that the organization's environment brings (Pettigrew, 1985, Pettigrew and Whipp, 1991). In the aviation sys-tem, a series of European projects (ADAMS, ADAMS 2, AMPOS) have analyzed the difficulties that organizations have in achieving effective change even in response to serious safety incidents.
Change is necessary, but it is risky. It is vital therefore that the reasons for organizational change failure are understood by those responsible for its enactment. Where change involves people in organizational processes it becomes primarily a 'human factors issue' and involves complex, multidimensional solutions. This is made difficult by the fact that change often results in "resistance" by employees and this requires careful management in terms of hearts and minds. It is necessary that managers who are tasked with implementing change initiatives understand clearly the complex individual, social and cultural dynamics that underlie the compliance or resistance to change on the part of personnel affected. For example, the idea that employees are resistant to change itself is contested by some researchers. According to Dent & Goldberg (1999), often individuals aren't really resisting the change itself, but they may be resisting the reduction in status, pay, or loss of comfort they perceive as being associated with it. They argue that, "it is time that we dispense with the phrase resistance to change and find a more useful and appropriate models for describing what the phrase has come to mean: employees are not wholeheartedly embracing a change that management wants to implement" (Dent & Goldberg 1999).
Strebel (1996) suggests that management should view how change looks from the employees' perspective, and to examine the terms of the "personal compacts"[9] currently in place. Unless managers engage with employees in redefining those terms, it is unrealistic to expect the workforce to fully buy into changes that alter the status quo.

Lessons Learnt from past experiences

In another EU funded project called HILAS one study identified a number of cultural dimensions associated with readiness to change:

[9] Personal compacts are defined by reciprocal obligations and mutual commitments that are both stated and implied in the relationship between employees and organizations.

- Organizational values reflecting information and communication openness, effective and valid human resource feedback mechanisms, and a focus on learning rather than punishment following mistakes.
- Organizational practices and policies that support vertical and horizontal interactions, participation and use of knowledge.
- Trust, which involves the credibility and justice of management initiatives (Cabrera, et al., 2009).

The HILAS project also demonstrated a number of case studies of successful change interventions, including:

- new strategic policies and programs to manage risk,
- diverse contributions to instituting a new corporate safety management infrastructure,
- the implementation of an effective process improvement program which fostered a new relationship between an airline and a maintenance company. (Ward & Gaynor 2009).

Through these case studies, HILAS confirmed the importance of system change as a key precursor of culture change, specifically, effective organizational processes which deliver beneficial outcomes to key stakeholders.

3 The Approach Proposed In MASCA

MASCA is concentrating on extending this state of the art by developing a dedicated Change Management System that will facilitate change initiatives through greater transparency, mentoring and information support in a sustainable manner and enhancing the commitment of organizational stakeholders toward its success. This will also involve training and competence development for change management skills and providing stakeholders with a better understanding of the various other aspects of the aviation system that their work impacts on and impacts on them.

Figure 1 illustrates the Process of Change pro-posed in MASCA. It is organized around three dimensions strategy (direction and goals), process (functional system) and competence (making sense of the system). Interpolated in the diagram are some other dimensions that are important in the management of change: the role of internal and external change agents; the extent to which the organization can act in a coherent integrated way; and the use of data to track change and validate strategy. The central motif of the diagram is that change is a process, from diagnosis of need to evaluation of outcome, which en-gages these different elements in different ways.

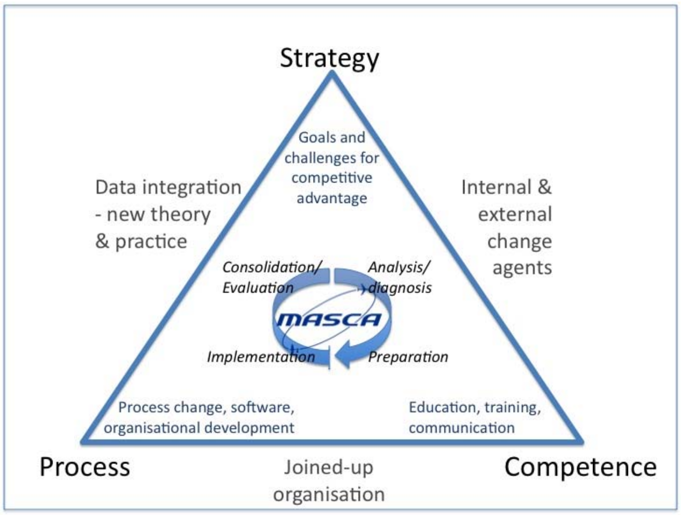


Figure 1. A schematic Representation of the change Process in MASCA

HF and Strategy

Strategy encapsulates the organization's relation-ship with its (often competitive) environment. The basic question is – what does the organization have to do to meet the foreseeable challenges in its environment? This may include relationships with its collaborators and competitors across an operational system (aviation system-of-systems) as well as across the lifecycle (design and manufacture, operate, maintain, regulate). The capacity to understand the competitive challenge alone is not sufficient. What is also needed is an internal intelligence on how the organization is performing, in order to understand how to change the operational system in order to match that challenge. This defines a gap that needs to be breached. That gap can be expressed in risk terms – the probability of adverse consequences if the anticipated gap is not breached. If the pro-posed changes to breach the gap can be modelled and assessed in terms of risk reduction, then the organization's strategy can be coherently organized in terms of identifying and prioritizing system risks and progressively working to manage and mitigate those risks. In strategic planning change is necessary, but it is risky. It is vital therefore that the reasons for organisational change failure are understood by those responsible for its enactment and that the views of the people involved are taken into account from an early stage in a more proactive manner.

System change

What are the minimal conditions for being able to purposefully change a functional system? It is necessary to know how that system works or functions as well as to be able to track what it is doing. This knowledge of functional activity needs to be sufficiently sophisticated to understand what needs to change in order to make the system work better. Finally it is necessary to know how to implement that change. Fulfilling these conditions gives leverage or power over the future performance of the system. The challenges outlined at the beginning of this argument were to be able to manage, maintain, regulate and design complex integrated systems in a global environment demanding them to be at the same time cheaper and safer to run, while delivering higher quality of service with lower environmental impact. It is also the case that the people staffing these systems want satisfying, sometimes fulfilling and challenging, work that respects their professional and personal dignity. If these challenges are to be addressed successfully then we have to fully meet all these criteria. This argument can be expressed as a set of methodologies built into a management tool.

How it works
Knowing how the system works involves modelling the relationships and dependencies (technical, social, informational) in the system. These govern the transformation of resource inputs through process activity into outputs. The modelling needs to be sufficiently sophisticated to project to a future sys-tem with appropriate changes.

What it is doing
The model also needs to be sufficiently well de-fined and grounded in the operation to be meaning-fully linked to real operational data measuring, con-currently, system inputs, activity and outputs. Different types of data, from everyday operational indices to in-depth analysis of rare events, should be integrated to provide knowledge of the relationships between normal and abnormal functioning of the system. These relationships define the risks inherent in the system and form the basis for estimating future risk following system change.

How to change
An objective of change is to reduce the risk the organization faces in meeting its strategic challenges, comparing current performance to projected demands. Reporting, investigation, audits, data gathering set off a sequence of management processes that manage and transform this information into actual system change. First they enable the systemic assessment of risk, the development of recommendations for action, implementation and the evaluation of the effectiveness of the actions taken.

The objective of risk assessment is to represent all analyzed risks in an integrated risk register that prioritizes the areas in which the organization has to take action to improve its systems and processes. Each incoming report or data point that exceeds some preset boundary receives an initial risk assessment to prioritize its status and identify urgent actions to be taken. One or many reports or data sets may then be combined to form 'projects' which rep-resent a common problem space. Three complementary types of risk analysis can then follow:
- Investigation & other qualitative analyses
- Analyzing probabilities of 'cause and effect' in multiple reports
- Analyzing probabilities of 'cause and effect' in operational data.

The objective is to construct and constantly up-date a synthetic picture of system risk from multiple sources of data that then prioritizes policies and pro-grams for change and redesign. Recommendations are derived using the process knowledge resource in the mapping tool to model the future process, conduct a future risk assessment and set evaluation metrics. The integration of modelling and data potentially provides the basis for simulation of future systems. Recommended actions can vary in their scope – local or systemic, technical or organizational.

The decision to implement change or redesign actions provides a new opportunity to form an implementation project from overlapping requirements (and this can generate fresh risk assessments). The trajectory of implementation will depend on the scope of the action, which is tracked through various stages.

Evaluation reconciles the projected benefits (and costs) derived from the recommendation phase with the actual benefits (risk reduction) and costs achieved in the implementation phase, taking into account the quality of the implementation. This closing of the loop depends on the quality of the initial systemic risk assessment that provides the baseline for the evaluation of change.

Competence
The argument about change highlighted two prevalent problems: the insufficient competence within organizations to manage complex change involving human, social and technical

systems; and the lack of a good practical theory to support professional and managerial practice in this area. This document is designed to begin to consolidate a us-able knowledge framework to address the latter. The issue is then to transfer this knowledge into practical know-how in the industrial organizations. This transfer of knowledge may have to happen at a number of different levels:

- An in-depth professional competence in the management cadre which can support the planning, direction and leadership of change and to fulfill specialist functions;
- More broad-based management training for middle management levels to ensure the facilitation of innovation and change both horizon-tally across the organisation's processes and vertically from senior management to the core operational staff;
- 'Knowledge transformation' activities to develop common understanding amongst diverse stakeholders in a change or innovation programme.
- General awareness programs for all staff

Education and training initiatives need to be sup-ported by effective communication. New social technologies can be used to stimulate wide-scale participation in achieving consensus about the objectives and process of change.

Organizational Culture

Organizational Culture has a dynamic role in maintaining system stability, from which it follows that a non-linear relationship to change may be typical. Qualitative change in collective understanding may only come after a cumulative aggregation of many minor shifts in the way a social group make sense of their situation; the shift may then be rapid and volatile; this is why the importance of consolidating and embedding change is so often emphasized. As culture cannot be directly 'managed' or controlled, attempts to do so often create an unofficial counter-culture (Kunda 1992). So if culture can-not be directly 'managed', equally we ignore it at our peril. But culture potentially includes everything that is thought about. Unless we think about how to focus on what is relevant, culture becomes 'what culture surveys measure'. The analysis of organizational culture can encompass a wide range of phenomena, including goals, vision and values; the perception of the system and its functions, often called climate; subcultures differentiated by roles and boundaries; the quality of engagement between people and organization, including dimensions like trust or alienation. It is therefore important to understand the variety of roles cultural analysis can have in organizational intervention.

Culture can be an index or criterion. Within the safety management literature having a good safety culture, a just culture, etc., often seems to be pro-pounded as a criterion of good safety management. For a cultural measure to perform this function effectively it has to be grounded in a demonstrated relationship to the functioning of the system it represents; the levels of knowledge and expectations need to be taken into account, as these can reverse predicted relationships. Also, as there are many layers of cultural meaning, what may seem as a qualitative value (e.g. a deficiency) at one level of analysis may have compensatory aspects at another layer of analysis; so great care needs to be taken in interpretation.

Culture can be seen as the active engagement of a collective in a process of change. Here the cultural analysis needs to be more agile, engaged and sensitive to local nuances that represent shifts of under-standing. The cultural analysis then can become part of the change process.

A classic approach to cultural analysis is to seek to establish fundamental meanings and values. Here the aim is to construct a rich in-depth interpretation from a broad range of material. While this may not be particularly useful from a short-term perspective in managing change, it can be extremely valuable in developing a strategic view of some of the challenges that need to be faced in changing an organization.

4 Current Developments in MASCA

Putting all these elements together makes it possible to start to plan a change management capability that can lead through the following stages (these form the central part of Figure 1):

- Establishing strategic goals
- Analysis and diagnosis of the problem, identification of the gap between current state and future challenge.
- Preparation for change
- Implementation programme
- Consolidation and embedding of change
- Evaluation

This argument represents a new attempt to resolve the contradictions between theory and practice – to provide a rigorous theory, which also provides a guide to action and intervention. The starting point is re-creating a functional analysis of socio-technical systems that forms the basis of a management tool that links operational modelling and analysis to the capability to assemble real operational data in a meaningful way, and a reporting system for managing risk and change.

During Phase 1 the research involved a systematic diagnosis of each of the four organizations based on semi-structured interviews supported by survey focusing on the key aspects of the MASCA framework (Strategy, Competence and Process).

The overall objectives during this phase of research included:

1. Clarification of the key strategic goals each of the companies are pursuing in order to meet the current business and economic challenges.
2. Understand and assess the current change initiatives each of the companies have in place to facilitate and ensure that overall, as well as local or departmental, strategic objectives are met.
3. Consolidate lessons learnt from previous change initiatives in each of the companies, including practical regulation and Human Factors related issues.
4. Assess each of the companies' readiness for change, for example a capability and maturity assessment.
    a. What is the level of current competence for change management?
    b. How is the current culture impeding and facilitating change?
5. Identify and analyze the key performance indicators and their underlying factors that each of the companies is currently utilizing in order to assess and measure if and to what extend their strategic goals are being met.
6. Link the MASCA change and evaluation programme to initiatives within each of the companies.

In order to meet the overall objectives and to achieve a much clearer notion of what companies need to do to implement change, the MASCA framework for assessing companies organisation's capability to change will be applied in each of the specific organizational case-studies.
The type of personnel involved for each one of the four companies during the field work was as follow:

- An average of 5-6 interviewees at Managerial level (Operation/ Production Manager, HR manager, Training and Quality Manager, Safety Manager, Financial Administrator, Technical support (IT) Manager, etc.)

- An average of 6 interviewees at Sharp end Operational level (maintenance technicians, handling agents (airports), flying and cabin crew (airline), shift supervisors, dispatch and planning agents, etc.)

Based on the analysis of this data in conjunction with industry based work-shops the following change case-studies have emerged:

1. the support to the organizational restructuring of a major Airport operator in Scandinavia in the area of strategic planning communication.
2. The definition of the process leading to the Safety Performance Indicators (SPI's) collection in an ICAO SMS within a major airline in EU and the support in the definition of the process leading to follow up on the consequent improvement loop
3. the support and training to be provided to the personnel of a small regional airline to plan and sustain a possible major change
4. Performance Management Framework for a small Airport southern Europe focusing on the changes needed for survival
5. The common objective of Collaborative Decision Making between airlines, airport operators, ground handlers and ATC, and the introduction of new technologies to support those operations.

Currently the MASCA research partners are working closely with the industry partners on the development of the implementation and evaluation trajectory of each of the above cases. In parallel to this activity the project has made considerable progress in identifying they key aspects of a change management 'tool-box' comprising of the following (see Figure 2)

- A tool to perform process modelling, risk analysis and monitoring (data reporting) building a data structure around the process itself.
- A Learning , training and Mentoring program, able to facilitate knowledge transformation workshops and to create common understanding amongst diverse stakeholders
- The use of simulation and serious games to align the training to the overall company strategies and test possible new processes and technologies.

5 The role of Training in MASCA

The overall MASCA programme has a primary focus on the transfer of change management capability into the organisations that are responsible for and involved in change. In order to effectively support the change initiatives a core component of the MASCA Change Management System is a framework for Learning, Training and Mentoring (LTM). The key elements of the LTM include:

- An on line Professional Masters Programme.
- Intensive & Focused Training Courses (both for internal and external change agents).
- Mentoring Framework (to support the competence and skills required for individual leaders which will be incorporated at the both the Masters level and at the level of the shorter training courses).

The summary of the results of the interviews carried out during the field work provided a preliminary assessment of the gaps where education and training could feed into. Table 2 below provides a high level overview of some of the specific areas of training required across the four companies in order to support effective change.

6 Conclusions and future developments

Successful change management involves careful planning, and the design of new operational and management processes, procedures and systems as well as their careful implementation.

Processes and system change both requires and results in cultural change. The key to success is arriving at the desired cultural climate as a result of the system changes introduced. That is why a careful understanding of people dynamics and needs is going to be kept central within MASCA in all the design and planning of the change initiatives. The focus on a specific training and Mentoring program is key to transferring change management capability into the organisations. Within MASCA the change management system need to be tested in supporting real change interventions involving the industrial partners and other end-users. In the next phases of research the team will be detailing each one of the change initiatives further in collaboration with a company internal set of MASCA change agents to identify the connected needed methods and support tools. This phase will be followed by a development phase preliminary to the actual change implementation. Implementation and evaluation are going to be planned and rolled out concurrently. Following evaluation the system will be revised and finalised against a plan for its exploitation across the aviation system and in other high risk systems.

Table 1 – High Level Overview of Specific Requirements across the four organisations

Major Scandinavian Airline	Airports operator	Small Regional Airport	Small regional Airline
Developing a change management strategy	Developing a change management strategy	Developing a change management strategy	Developing a change management strategy
Leading Change/Building a Change Team	Leading Change/Building a Change Team	Leading Change/Building a Change Team	Leading Change/Building a Change Team
Leadership/Mentoring for Managers.	Leadership/Mentoring for Managers	Leadership/Mentoring for Managers	Leadership/Mentoring for Managers
Link Leadership training to operational & strategic challenges		Improve Team Working at ops level (clear roles & responsibilities etc)	
		Supervisory Training/Facilitation	Supervisory/Mgmt Training/Facilitation
	Communication (from top-level to ops)	Communication (from top-level to ops)	Communication (from top-level to ops)
		Performance Appraisals	Assess the Trainers
e-learning	e-learning	e-learning	e-learning
Basic Human Factors training in context	Basic Human Factors training in context	Basic Human Factors training in context	Basic Human Factors training in context
Decision Making	Decision Making	Decision Making	Decision Making
Link training to other CI projects (e.g., Lean)	Link training to other CI projects	Link training to other CI projects	Link training to other CI projects
Resistance Management Plan (overcoming change fatigue)	Resistance Management Plan (overcoming change fatigue)	Resistance Management Plan (overcoming change fatigue)	Resistance Management Plan (overcoming change fatigue)
MASCA CMS	MASCA CMS	MASCA CMS	MASCA CMS

Acknowledgements

The above mentioned research has received funding from the European Commission's Seventh Framework Programme FP7/2007-2013 under grant agreement 266423 “MASCA”.

References

ACARE (2001) European Aeronautics: A Vision For 2020 Report Of The Group Of Personalities January 2001. Available on the website http://www.acare4europe.org/docs/Vision%202020.pdf [last accessed May 2011].

ADAMS (1994 - 1999) Human Factors in Aircraft Dispatch and Maintenance. Under the EU BRITE-EURAM Programme. Contract No. BRPR-CT95-0038, Project No: BE95-1732. Brussels: European Commission.

ADAMS2 (2001 - 2004) Human Factors in Aircraft Dispatch and Maintenance. Under the EU BRITE-EURAM III Programme for Industrial Materials and Technology. Project No: GRD1-2000-25751. Brussels: European Commission.

AMPOS (1999 – 2001) Aircraft Maintenance Procedure Optimisation System. Project No: 29053. Brussels: European Commission.

Cabrera, D., Ramos Sapena, Y. & Villegas Velásquez, O. (2008) Cultural Implications of Implementation Requirements. Deliverable to the EU Commission as part of the HILAS Project. IP8. Proposal/Contract No: 516181. Brussels: European Commission.

Dent, E. & Goldberg, S. (1999, March). Challenging "resistance to change." Journal of Applied Behavioral Science 25-41.

ICAO (2009) Safety Management Manual (SMM) Doc 9859 AN/474.

Kotter, J. P. (1995, Mar-Apr). Leading change: Why transformation efforts fail. Harvard Business Review, Vol. 73, No. 2, March/April, pps. 59-67.

Maurer, R. (1996). Beyond the wall of resistance: Un-conventional strategies that build support for change. Austin, TX: Bard Books, Inc.

Pascale, R., Millemann, M., and Gioja, L. (1997). Changing the way we change. Harvard Business Review, Vol. 75, No. 6, November/December, pps. 127-139.

Pettigrew, A. (1985). The Awakening Giant: Continuity and Change at ICI. Oxford: Basil Blackwell.

Pettigrew, A. and Whipp R. (1991). Managing change for competitive success. Oxford: Basil Blackwell.

Porras, J. I. and Robertson, P. J. (1983). Organization development: theory, practice, and research. In M. D. Dunnette& L. M. Hough (Eds.).The handbook of industrial and organizational psychology, 3, (pp. 719-822). Palo Alto, CA: Consulting Psychologists Press.

Robbins and Gioia Survey 2001 as found on the web page http://www.it-cortex.com/Stat_Failure_Rate.htm accessed on January 2010.

Strebel, P. (1996, May/Jun). Why do employees resist change? Harvard Business Review 86-92.

Ward M, Gaynor D. 2009 The Ergonomic Design and Implementation of an Improvement Initiative for an Aircraft Base Maintenance Check Proceedings of the Irish Ergonomics Society Annual Conference 2009 Edited by Leonard W. O’Sullivan ISSN 1649-2102.

OPERATIONAL ANALYSIS/VALIDATION OF FUTURE COCKPIT TECHNOLOGIES TO SUPPORT ALL CONDITIONS OPERATIONS

Joan Cahill, Rabea Morrison and Nick Mc Donald

Aerospace Psychology Research Group (APRG),
School of Psychology,
Trinity College Dublin
Dublin

Abstract
This paper reports on human factors/human computer interaction (HCI) research concerning the operational validation of future cockpit concepts, designed to address the problem of all conditions operations, as part of the All Condition Operations and Innovative Cockpit Infrastructure (Alicia) project. Specifically it reports on a subset of this research, namely the design of a questionnaire used (1) to envision in more detail the nature of the proposed cockpit technologies, and (2) to assess whether the new technology concepts address existing operational problems and deliver on key operational requirements. The purpose of this paper is to demonstrate the requirement for HCI research to go beyond issues of task and user interface design, so as to properly address issues of operability. It is argued the typical remit of operational assessment (i.e. the assessment of task workflows, task workload and operator situation awareness), must be supplemented by a broader analysis of 'operational' issues. This includes the design of operational processes/procedures, teamwork, system information flow, operational benefits and implementation barriers.

Introduction

Introduction
This paper reports on human factors research concerning the operational validation of future cockpit concepts, as part of the All Condition Operations and Innovative Cockpit Infrastructure (Alicia) project. Primarily, this paper focuses on one particular strand of this research. This concerns questionnaire based research, designed to both specify and evaluate the proposed cockpit technology solutions from an operational perspective.

First a background to this research is provided. A summary of the overall Human Computer Interaction (HCI) plan for the project, along with the TCD research role, is then provided. Following this, the collective operational analysis methods used in this project are presented. The specific methodological approach for the questionnaire based research is then outlined. High level research findings are then reviewed. This is followed by a discussion of research methods. Lastly, some preliminary conclusions are proposed.

Background
HCI research provides a direct input to new product development. Critically, it serves to identify and evaluate the requirements for new operational concepts/procedures, and the associated technology requirements. This role can be summarized in relation to the mantra 'design for operability'. Accordingly, HCI research must ensure that new technology concepts address existing operational problems and deliver on key operational requirements at a process, task, and information flow level.

On this subject, a number of research observations are worth noting. Through our participation in several recent EU 6th and 7th framework project, the Aerospace Psychology Research Group (APRG) at Trinity College Dublin (TCD) has learned several important lessons in relation to technology development and process/task analysis and redesign. Technology development is not simply a matter of availability. New technology should be introduced to solve real operational problems. To do this, technology development needs to be driven by an understanding of task and process activity (Mc Donald and Morrison, 2006, Morrison, 2009). However, standard operating procedures (SOP) do not always provide an adequate description of task and process activity (McDonald, Corrigan and Ward, 2002). Often, SOP neglect to describe actual social processes such as informal work practices, and communication/co-ordination processes (McDonald, Corrigan and Ward, 2002, Ward, 2005, Mc Donald and Morrison, 2006, Morrison, 2009, Cahill, 2010).

Further, existing HCI methods fail to address the broader operational issues underpinning the envisionment of new technologies (Cahill, 2011, 2010, 2008). Critically, existing methods fail to address the relationship between process, task and tools/information flow, and the requirement to advance new technologies which solve operational problems at a process, task and information flow level. As demonstrated by Cahill in prior research (2011, 2010, 2008), both process modelling and collaborative design and evaluation activities are central to new process, task and technology envisionment.

Given the above, HCI research must start from a broad definition of 'operability'. In addition to focusing on issues related to task design, situation awareness and workload, we must also consider certain wider issues such as process design, team concepts and system information flow. Further, we must understand the impact of new technology on the overall operational system. In so doing, the operational cost benefit must be assessed. This spans issues pertaining to process re-design, resources/workload, organizational capacity and training (Morrison, 2009). Thus, in this definition, HCI research is not simply about task and user interface design (Cahill, 2010).

These requirements in turn set an agenda for HCI research. Specifically, it creates a requirement for HCI researchers to collaborate with technology partners in relation to the specification of new operational concepts and allied task support technologies. This in turn, directly contributes to the definition of user requirements, system requirements and HCI requirements. Moreover, it provides a baseline for HCI evaluation processes.

Project Problem
Vision 2020 (Group of Personalities, 2001) proposed a goal of an improved Air Traffic Management (ATM) system that can manage up to three times more aircraft movements than today, using new operational concepts and systems that permit aircraft to operate in all weather conditions, to fly closer together at lower risk, and to run on schedule 99% of the time (Alicia Consortium, 2009). Poor weather is one of the most disruptive factors in European aviation today. Weather factors can be classified into four main categories. This includes: ice, wind, rain, and rapidly changing weather conditions. As indicated in Figure 1, these have an impact at different points in the flight timeline. Specifically, at relevant points in the flight timeline, weather factors/problems have an impact at (1) an operator level (i.e.

Pilot task level) and (2) an operations/process level (i.e. management of single and multiple flights by ATM services and Airlines/Flight Operations). At a task level, weather problems result in increased task complexity and workload. As the task situation becomes more difficult, situation awareness and error management is essential to the avoidance of adverse outcomes (i.e. safety critical incidents/accidents). As illustrated in Figure 2, from an operations/process level, weather problems have an impact on the management of both single and multiple flights. Often such problems result in delays, diversions, cancellations and runway closures. For instance, in the SESAR definition phase, it was estimated that 16,800 airline flights were cancelled in 2007 in Europe due to low visibility conditions, and in some major airports almost 50% of arrival delays are due to low cloud and poor visibility (Alicia Consortium, 2009). Although current systems can support approach and landing in low visibility (for suitably equipped runways and aircraft), if aircraft cannot taxi at the same rate as in good visibility, then the time between take-offs and landings will inevitably increase and lead to delays at the busier airports (Alicia Consortium, 2009). In addition, low visibility is a major contributing factor to runway incursions: currently there is a runway incursion every three or four days in Europe and a near collision due to runway incursion every two or three months (Alicia Consortium, 2009).

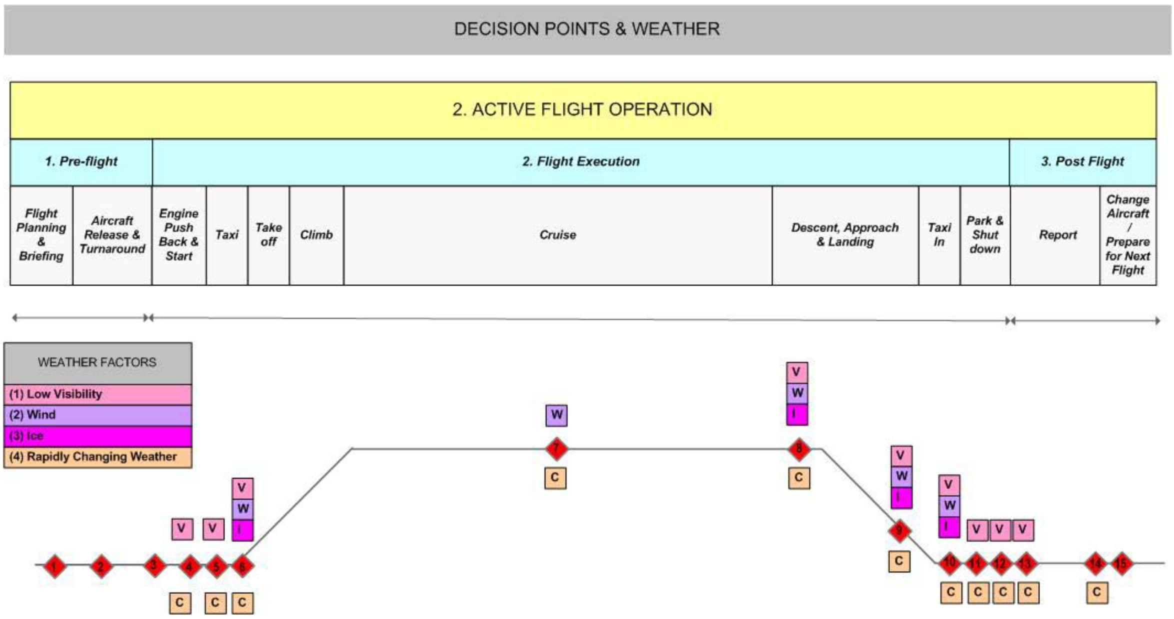


Figure 1: Weather Factors and Location in Flight Operations Process

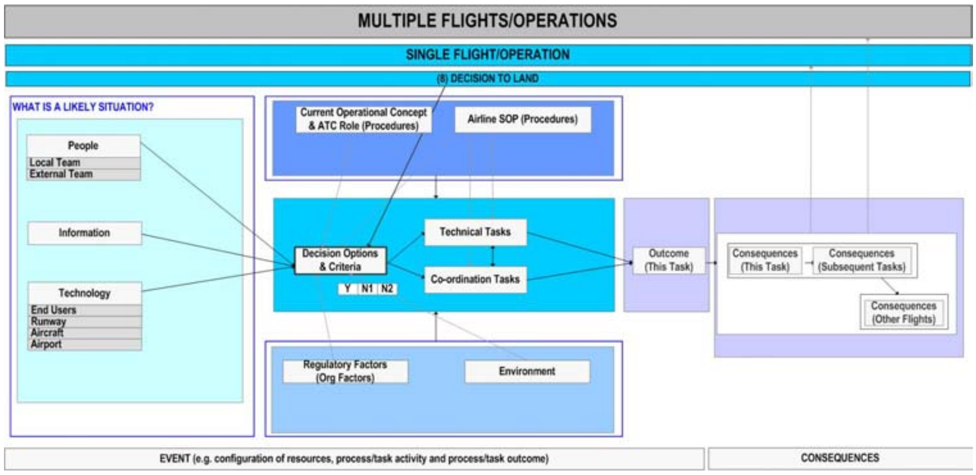


Figure 2: Weather Problems and Impact on Flight Operations Process (Single Flights and Multiple Flights)

In an effort to overcome some of these problems, new flight operations concepts along with the respective ATC and Flight Crew technology solutions, have been advanced. This includes the Single European Sky (Eurocontrol, 2009), Free Flight (Johnson, Battiste and Bochow 1999), and the movement from a two crew cockpit to single crew or fully automated flight. The Free Flight Concept is of particular interest. A critical feature of this new operational concept is the proposal to share flight information across all relevant stakeholders (i.e. ATC, Flight Crew, Airport Operator and Ground Operations). To date, technology development has focused on implementing new ground/air data links – supporting 4 D trajectory management and assisting information sharing between Flight Crew and ATC during the flight execution phase. This research is being advanced as part of the SESAR project (Derisson 2007). However, the specific issue of enhancing cockpit weather management, to support the ACARE vision, is not directly addressed in this project. Nonetheless, there is much to learn from these initial concepts. The advancement of new technologies which enhances information sharing across different stakeholders (both directly and indirectly) is a good idea. Potentially, these concepts might be further exploited, to take into account information sharing between Flight Crew and certain other operational functions currently not addressed in the SESAR concept (i.e. Met, Airline Flight Operations, Maintenance and so forth).

Introduction to Alicia Project

The ALICIA project is 7th framework EU project which addresses the ACARE objective of increasing time efficiency within the future air transport system, whilst maintaining safety. The aim is to develop new and scalable cockpit architecture and technology platform which will extend aircraft operations in degraded conditions and in particular, reduce weather-related delays by 20%. Technologies will be developed for both rotary and fixed wing (business jets, regional jets and large wide body aircraft) platforms. The consortium comprises aircraft manufacturers, avionics/technology providers, human factors/human computer interaction researchers, and a panel of consultant Pilots.

Before explaining the overall HCI research plan for this project, and the specific contribution of this operational analysis/validation research, several project challenges must be noted. Firstly, although the project team includes a small panel of consultant Pilots, actual operators (i.e. Airlines, Airport Operators, MET and so forth) are not directly involved in the project. With this in mind, the Alicia project consults with an external panel of operational experts on a biannual basis, to learn from their experiences and validate the proposed operational and technology concepts/solutions. This group is termed the Alicia External Expert Advisory Group (EEAG) and comprises expertise in several domains including air traffic management

(ATM), flight operations, search and rescue, meteorology, flight safety and cockpit technology certification processes. Thus, it is expedient that HCI researchers optimize collaboration with both internal (i.e. consultant Pilots) and external (i.e. EEAG) project 'resources', to ensure that end user feedback/validation is obtained.

Secondly, the project consortium brings several diverse technologies suitable for both platforms (i.e. both rotary and fixed wing platforms) to the table. Each of these technologies is at different levels of maturity (i.e. from paper concept to low fidelity prototype demonstration). There is a strategic interest in advancing these technologies, in the context of the Alicia project brief/goal. As such, HCI researchers must steer the technology development, to ensure that candidate technologies properly address operational problems (i.e. that is, not simply developing technology because it is available).

Lastly, operational value will be a key measure of project success. The Alicia project goals as defined earlier are sub-divided into project objectives, evaluation categories and evaluation objectives. There are four evaluation categories; namely (1) all conditions operations, (2) new cockpit concept, (3), synthetic environments and (4) usability. Critically, key performance indicators (KPI's) are associated with each evaluation category. Thus, future technology evaluation will necessarily involve justifying the technology from an operational perspective and specifically, against key all conditions operations and usability KPI's. Consequently, HCI research must ensure that evaluation activities address this issue.

Overview of HCI Research Plan and APRG/TCD Role

Overall HCI Research Plan (All Partners)

The overall HCI research plan involves several phases of research. In the first phase, this includes (1) an initial operational analysis, (2) task analysis and (3) an initial specification of high level operational concepts and technology requirements. In the second phase, a subset of the technologies will be selected for development. Three stages of prototyping and evaluation are planned. These are defined as Level 1, 2 and 3 prototyping. In the first stage, the high level HCI requirements for each technology will be defined and low fidelity prototypes produced and evaluated. In the second and third stages, these prototypes will be further specified and evaluated. Finally, the third phase of research will involve more in depth simulations/evaluations of the proposed technologies. This may involve the use of synthetic environments.

Alicia Co-ordination Activities & Research Clusters

To facilitate interaction/co-ordination between the different technology partners and between technology partners and HCI researchers, a small number of clusters have been formed comprising partners with related technologies. For example, synthetic vision, touch-screens and input devices. Each technology cluster has a leader who co-ordinates among cluster members and across clusters. Further, three separate clusters have been set up to co-ordinate research related to (1) weather, (2) operational processes and scenarios, and (3) evaluation activities.

TCD Role and HCI Research Objective

The focus of our research in the Alicia project is on fixed wing operations. The principle goal is to support the operational validation of the proposed ALICIA technologies. As such, our research endeavours to both pose and answer several 'operational' questions. For example: (1) what are the existing operational problems?, (2) how can we steer the technology/HCI research so that it properly address existing operational problems?, (3) what are the specific HCI requirements for these technologies in relation solving operational problems at different levels (i.e. process, task and information flow level)?, (4) do these technologies/prototypes solve operational problems, and if so, how?, and (5) what are the specific information flow requirements of the overall cockpit solution in relation to the broader socio-technical picture?

Overview of Collective Research to Date

The methodological approach adopted in this research is underpinned by the Knowledge Space Model (KSM) – an operational analysis framework, advanced by researchers at the APRG (Mc Donald and Morrison, 2006, McDonald 2006, Ward, 2005, Baranzini, 2009, Morrison, 2009). An illustration of this framework is depictured in Figure 3 below.

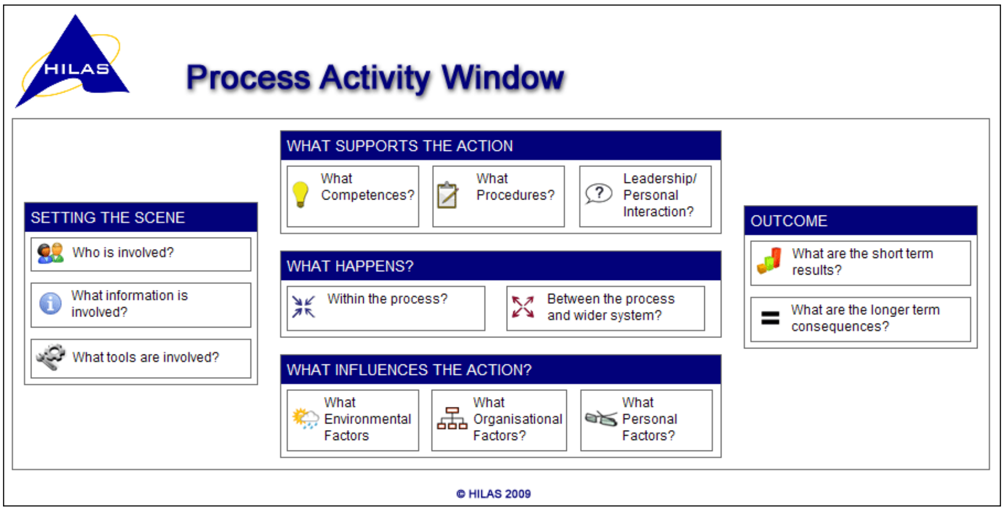


Figure 3: Knowledge Space Model (KSM) Framework

The specific operational analysis methods deployed in this project have been designed to fit to the project schedule for HCI activities, along with answering the operational analysis questions detailed earlier, while bearing in mind the different project challenges (i.e. access to operators/end users, strategic technology agenda, and justification of research against KPI's). To date, two phases of research, each spanning several strands of enquiry have been undertaken. These are summarised in Table One below.

Table One: Overview of TCD Operational Analysis Research (All)

OVERVIEW OF OPERATIONAL ANALYSIS/VALIATION RESEARCH (ALL)	
PHASE 1 (Year 1)	
1	Advancement of KSM framework in project context
2	Literature review: weather and operational problems/requirements
3	Analysis of operational problems and specification of operational requirements
4	Specification of high level Fixed wing scenarios and associated risk assessment
5	High Level process mapping (current process)
PHASE 2 (Year 2)	
1	Detailed process mapping (current process)
2	Specification of detailed weather scenarios
3	Specification of Key Performance Indicators (KPI's)
4	Operational analysis/validation of candidate technologies (i.e. extended questionnaire research)
5	Specification of new planning concept
6	Input to specification of technology requirements

7	Input to evaluation activities

Collectively, these are linked up activities. Further, it is anticipated that there will be several iterations of the above research, through the course of the project. Moreover, linking to these two phases of research, a number of collaborative research activities have been undertaken with other technology and research partners. This includes:

- Review and evaluation of Level 1 prototypes (i.e. technology concepts)
- Participation in a series of meetings with EEAG – to define operational requirements and technology concepts
- Participation in cluster groups – to support the specification of new technology requirements (level 2 prototyping)
- Specification of common evaluation framework for level 2 prototyping

Overall, this research (see Table 1) spans several phases in the HCI design process. This includes (1) problem specification, (2) the specification of high level operational problems and requirements, (3) technology concept envisionment/definition (4) detailed specification of user requirements, (5) detailed specification of HCI requirements, (6) design/prototyping and (7) evaluation.

Operational Analysis/Validation of Future Cockpit Concepts (Questionnaire)

Introduction

As indicated in Table One, the questionnaire based operational analysis/validation research has been undertaken in year two of the project, and in parallel to other phase 2 activities – in particular, Level 2 prototyping and evaluation. The overall objective of this piece of research is to:

- Support technology envisionment from the perspective of the prior analysis of operational problems and requirements
- Support specification of technology requirements - in particular in relation to teamwork/co-ordination, information flow and HCI design
- Elucidate the operational benefits of the proposed technologies
- Examine potential implementation issues/barriers
- Support broader technology evaluation activities (i.e. other Level 2 evaluations undertaken by project partners)

The technologies to be evaluated include: Heads up Displays (HUDS), Touch-screen, a Weather & Flight Planning Application, Dockable Displays, Equivalent Vision Systems, Conformal Symbology, Minimized Over-head Panel, Cursor Control Devices & Side Sticks, Peripheral Vision Horizontal Display, and Large Area Head-Down Displays.

Overview of Research

As defined in Table Two, this research involves several phases of data gathering, data analysis and communication/review of findings. In terms of the overall logic, each candidate technology is first evaluated separately. Following this, the total cockpit technology picture is analyzed.

Table Two: Overview of Questionnaire based Operational Analysis/Validation Research

OVERVIEW OF QUESTIONNAIRE RESEARCH	
BACKGROUND ANALYSIS	
1	Specification of operational analysis framework and associated operational questions
2	Review of project documentation/reports and produce draft/provisional answers to operational questions
SINGLE TECHNOLOGY PICTURE	
3	Initial communications with partners to explain approach
4	Data gathering: email technology specific questionnaire to relevant partners
5	Data analysis: review and integration of findings/feedback
6	Feedback to partners 1:
7	Additional specification and review (iterative...)
8	Elicit feedback of EEAG 1
9	Additional specification and review (iterative...)
10	Feedback to partners 2:
11	Overall report (evaluation of individual technologies and specification of requirements)
TOTAL COCKPIT PICTURE	
12	Evaluation of total Cockpit Technology Picture
13	Specification of technology integration requirements
14	Feedback to partners 3
15	Elicit feedback of EEAG 2
16	Final report: Final evaluation of technologies and specification of requirements (both individual and total cockpit level)

The electronic questionnaire comprises (1) background information about the purpose and method and, (2) a series of open questions and sample answers. All questions and answers (albeit provisional) are classified in relation to L1 and L2 prototyping. The questions are structured in terms of a series of topics following from the KSM framework. Further, the questions address specific weather management and cockpit solution user interface design issues. Overall, questions are classified as follows:

- Operations and Alicia Relevance
- Process and decision points
- Task
- Information
- Information: Teamwork and Co-ordination
- Information: HCI Design and Interaction
- Information & Weather Management
- Implementation
- Benefits and Impact

The provisional answers provided to partners were based on a review and analysis of relevant project documentation and research outputs. This includes specific L1 prototyping technology documentation, feedback from a L1 prototyping technology questionnaire circulated previously by the evaluation cluster, feedback provided by the consultant Pilots at project meetings, and feedback provided from operational experts at prior EEAG meetings. The answers provided by TCD to specific operational questions are provisional/draft. In certain cases, there are gaps (i.e. no answers provided). It should be noted that the whole point of providing initial/draft answers, is to get partners thinking about the operational dimensions of these technologies, and to promote a discussion/review of these technologies from an operational perspective.

Before circulating the electronic questionnaires to cluster leaders, the researcher engaged in a series of communication sessions with the Work Package (WP) leader, the panel of consultant Pilots, and with members of the evaluation cluster, to explain the overall approach adopted, and how it links with the broader HCI researcher activities of the project group. This was mediated through a series of teleconferences, and followed from prior communication activities (i.e. meeting presentations and research outputs) in the first year of project research. Further, the researcher emailed relevant partners with background documentation linking to relevant phase 1 research outputs and parallel phase 2 research outputs. This included process maps, example scenarios and KSM framework information.

The electronic questionnaire was then emailed to partners along with a summary explanation. Further, a follow up telephone call was made to all cluster leaders to further explain the approach, to clarify our research expectations, and to answer any questions. Specifically, a total of nine technology specific questionnaires have been sent to eighteen cluster members. This includes nine cluster leaders, and a further nine cluster members. Further, all nine technology specific questionnaires have been sent to each of the consultant Pilots (i.e. each of the 4 pilots providing individual feedback about specific candidate technologies). The questionnaires were reviewed by all participants and returned in Microsoft review/mark up format. Following this, the collective feedback was compiled, integrated and analyzed.

Current Status of Research & Next Steps
The review and integration of questionnaire findings is an ongoing process. At present, a preliminary report is being compiled which integrates the above feedback for each of the candidate technologies. Further, a draft report is being produced, providing general information about the operational validation of each of the technologies, the relationship between technologies and KPI's, and technology specific HCI design requirements. This corresponds to Phase 6 as detailed in Table Two above. Following from this, a series of co-ordination/communication activities is planned with wider cluster members and with the EEAG. Once this is complete, the total technology picture will be evaluated. As outlined in Table Two, this too will involve participatory problem solving and evaluative activities with cluster partners, consultant Pilots and members of the EEAG.

High Level Findings

Flight Operations Process and Decision Points
As indicated in Figure 4 below, a flight operation can be conceptualized in relation to a series of process phases, process sub-phases and decision points.

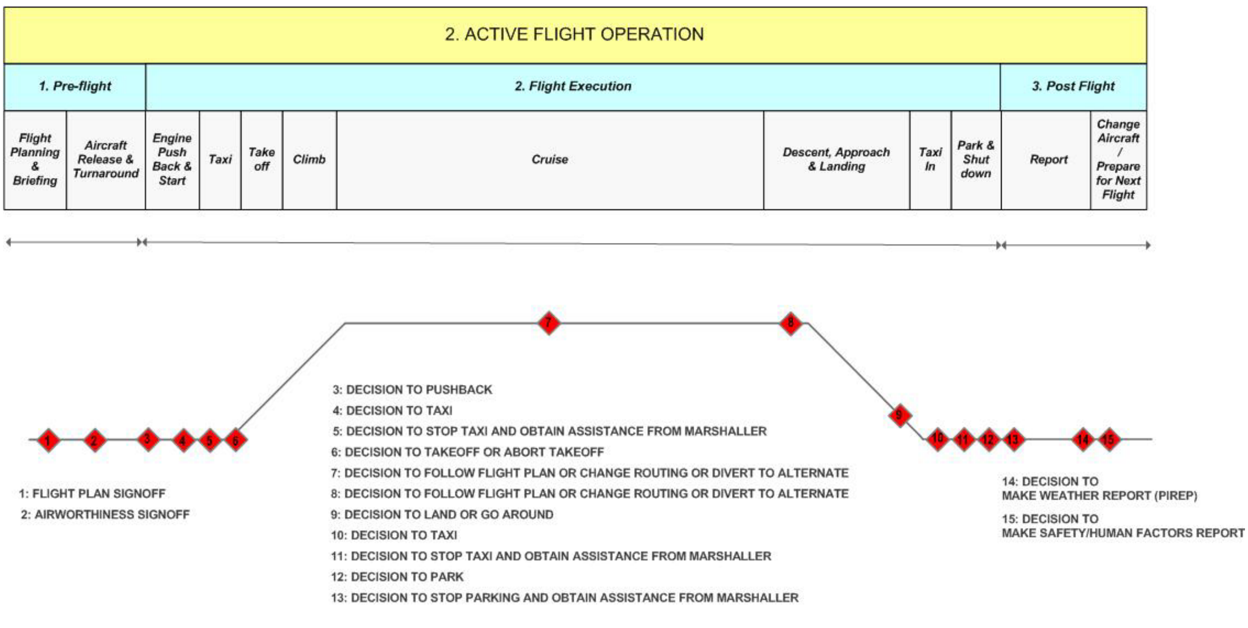


Figure 4: Flight Operations Process and Decision Points

Each flight involves information sharing and co-ordination across many different actors involved in a range of operational processes, which converge at the specific decision points. This includes the following actors:

- ATM
- Airline/Flight Crew
- Airline/Flight Operations Control
- Airline/Safety
- Weather Services
- Ground Operations
- Maintenance
- Airport Operator

Both the specific (1) operator and (2) collective team information requirements for each decision point can also be specified. Currently, there are many gaps in relation to information sharing across relevant actors at these decision points. In particular, there are gaps in terms of the provision of decision support information to Flight Crew to support weather management and flight safety. Critically, there is a requirement to enhance information sharing across these different actors at key decision points, to facilitate common situation awareness and team based decision making.

New Operational concepts

New operational concepts are required, to meet the relevant ACARE targets (Group of Personalities, 2001), and to solve existing operational problems. This includes enabling operations in poor weather conditions, enabling operations at airfields with reduced equipment and/or support (i.e. ATM control) and potentially, single crew operations. This is likely to be underpinned by new planning concepts. It is likely that this will necessitate extensions to existing SESAR planning concepts (SESAR Consortium, 2008), to assist information sharing across all relevant actors in the process. This spans the spectrum of actors involved in the flight planning process and the active flight operation.

Technology Opportunities

New technologies offer ample opportunity to solve many of the operational problems identified in the first phase of research. At a cockpit/task level – this includes problems relating to (1) the management of information, (2) access to information, (3) workload, (4) situation awareness and (5) error avoidance and management. At a team level, this includes the sharing of information across the network, enhancing the existing concepts of information sharing as defined in the SESAR SWIM concept (SESAR Consortium, 2008, SESAR Joint Undertaking, 2009). Lastly, at a process level, it is envisaged that a subset of the technologies (such as the HUD technology) will directly facilitate operations in low visibility or poor weather conditions, which in turn will have an impact on operational capacity (i.e. management of delays, cancellations and so forth).

Key Cockpit Technology Requirements

The key technology requirements are summarized in Table 3 below.

Table 3: Key Technology Requirements

KEY TECHNOLOGY REQUIREMENTS	
1	Enabling real time or close to real time information sharing across different actors at key decision points in flight timeline
2	Improved access to information
3	Faster and more intuitive access to information
4	Improved presentation format for flight information (including flight symbols, 3D, 4D etc)
5	Core flight control information presented in heads up format
6	Improve communication and presentation of weather information (i.e. 4 D weather and enhanced symbology)
7	Provision of predictive information (i.e. flight planning and weather management information)
8	Provision of decision support information in relation to the management of weather issues
9	Improved cockpit information integration

Discussion

Design for Operability

In order to assess the ability of the ALICIA technologies to fulfill the capability to realize the ALICIA concept, it helps to have a full and systemic picture of the operational situation in which the technologies are to be deployed. After all, it is the operational process itself, enabled by technology, which delivers the enhanced outputs that need to be demonstrated. The relevant technologies produce and manage information. We need to understand how that information can be leveraged in the operational process, supplementing or transforming task activity and co-ordination between people, to achieve an enhanced operational outcome.

HCI Methods & Scope of HCI Research

As demonstrated, the range of methods utilized in this research bridge existing gaps in the HCI design process, in relation to envisioning and validating the candidate technologies from a broader operational perspective. In particular the operational analysis/validation questionnaire has been used (1) to envision in more detail the nature of the proposed technologies, and (2) to assess whether the new technology concepts address existing operational problems and deliver on key operational requirements, at different levels (i.e. process, task and information flow levels). Thus conceived, HCI research must go beyond task analysis and user interface design. Specifically, it must evaluate the role of technology in the broader socio-technical system, and as part of this, design for operability.

Collaborative Research

Our research task was to find a mechanism in which to steer technology development from an operational perspective, and to gather information/feedback about the operational value of the proposed solutions. Working with end users/operators (i.e. end users of technology and/or those actors who benefit from technology) is central to this. To this end, we are working closely with internal (i.e. consultant Pilots) and external (i.e. EEAG) experts. Further, the various technology partners have Human Factors/HCI experts along with staff with relevant domain experience (i.e. cockpit and ATM operations), whose expertise can be utilized. Critically, there is a requirement to extract and co-ordinate the knowledge of these different people, to ensure that the technology specification addresses operational problems and provides value from an operational perspective.

This research has involved a learning curve for other HCI researchers working in the project, technology partners and internal/external operational experts. Thus, it was essential that we explain our approach and situate it in the context of the overall HCI process and broader project HCI research plans. In this way, conversations/presentations at project meetings, teleconferences with project partners, and follow up telephone calls with cluster leaders, proved invaluable in obtaining buy-in and comprehension from partners. Further, the circulation of process mapping documentation, operational scenarios and background KSM information provided necessary back-up reference information.

It could be argued that the provision of initial answers to questionnaire questions might bias the results/feedback. However, this is justified on a number of grounds. Initial discussions with technology partners indicated that they might not be in a position to answer operational questions on their own. This was attributed both to the maturity of the L1/L2 concepts/prototype, and to their lack of familiarity with this form of analysis. Further, the internal experts (i.e. panel of Pilots) noted that they did not have deep knowledge of the specific technologies. Thus, some form of collective collaboration was required. Therefore, the provision of draft answers enabled further envisionment and problem solving, on the part of both the technology partners and Pilot panel. In this way, the questionnaire formed the basis of a 'midwife' approach, and as such, departs from formal questionnaires (i.e. which merely pose questions).

Evidently, the results reflect the personal/expert view of the different partners. To date, the integration of feedback across different partners has not been straightforward. In a small number of cases, partners have expressed conflicting views regarding the operational value of specific technologies, interaction methods, and implementation issues. In this respect, it should be noted that this methodology (see Table 2) adopts a cyclical/iterative approach. Various cycles of data gathering, analysis and communication/dissemination are envisaged. It is anticipated that this will facilitate consensus across project partners, as to the particular nature of the technology solutions and their operational merits.

Proposed Methodology and Participatory Design Methods

Overall, this research can be characterized as adopting a 'participatory' approach. Arguably, the focus on envisioning the candidate cockpit technologies in relation to operational scenarios, links to the scenario based design approach defined by Carroll (2000). That said it extends this approach insofar as it addresses scenarios in relation to the broader socio-technical system and issues of operability, as opposed to focusing on individual actors and their associated task workflows. Further, insofar as the questionnaire supports envisionment activities, it can be linked to the envisionment and collaborative design activities defined by Muller (2003) and Bødker and Burr (2002). However, it should be noted that this approach does not involve actual prototyping. The electronic questionnaire is text based, and does not invite participants to draw out future concepts. Moreover, envisionment activities are process centric and span different conceptual/design levels (i.e. process, task and system information flow). Therefore, one of the challenges of this approach will be to integrate it with ongoing design activities (i.e. match between design at conceptual level and design at practical level).

Research Outputs/Benefits

To date, the questionnaire has proved useful in contributing to the specification of future operational concepts and the allied specification and evaluation of new cockpit technologies. At an individual technology level, this feedback has been used to:

- Locate individual technologies in the operation/process (where used in process and by whom)
- Identify what operational problems the technology addresses and how it addresses these problems
- Identify high level user requirements which the proposed technology delivers on
- Identify specific HCI requirements in relation to task workflows, user interaction, information flow, and the presentation of information
- Understand benefits and impact of technology
- Understand potential technology barriers

It is anticipated that this will also yield outputs in relation to the total cockpit information integration picture. Moreover, at a broader project level, this research provides a means to assess the value of each of the technologies in relation to project KPI's – in particular, KPI's in the all conditions operations category and the general usability category.

Conclusions

The HCI methods utilized in this project address the operational analysis/validation of future cockpit operational concepts and associated technologies. This spans several methods including the analysis of operational problems, process mapping, scenario specification and the analysis/validation of proposed technologies from an operational perspective (i.e. questionnaire based research). The specific operational analysis/validation based research (i.e. questionnaire) goes beyond the typical remit of operational assessment (i.e. the assessment of workflow, workload and situation awareness). Importantly, it considers how the technology addresses operational problems, the impact of the proposed technology on existing operational processes and process redesign requirements, the different actors in the process and team information sharing requirements, particular technology HCI requirements, operational benefits and implementation barriers. As shown, this approach spans various phases in the user centred design lifecycle, from problem identification, to concept envisionment, to HCI specification and evaluation. Critically, the research methodology addresses project challenges related to obtaining access to the operational community, and the plan/timescale for wider project research. Further, it directly addresses the key operational questions necessary to validating future cockpit concepts. Overall, this is an iterative approach. Various cycles of data gathering, analysis and communication/dissemination are envisaged. To date, questionnaire feedback indicates that the proposed Alicia technologies provide a strong opportunity to solve existing operational problems at different levels (i.e. process, task, system information flow etc), and in so doing, deliver on ACARE objectives.

Acknowledgements

The authors would like to thank the European Commission for sponsoring this research. Further, we would like to thank the Alicia Project Consortium for their support – in particular, the project co-coordinator, Westland Helicopters Ltd.

References

All Condition Operations and Innovative Cockpit Infrastructure (Alicia) Project Consortium. (2009). *Annex 1: Description of Work*. Project Number 233682. Funded by European Commission - Seventh Framework Programme – Theme 7: Transport (Including Aeronautics).

Baranzini, D. (2009). *Team Concepts in Maintenance*. Unpublished doctoral dissertation, Trinity College Dublin, Dublin, Ireland.

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

Cahill J., Mc Donald N., Losa G. (2011). 'Intelligent Planning and the Design of a New Risk Based, Intelligent Flight Plan'. *Cognition, Technology and Work.* Volume 13. Number 1, pp 43-66, Springer, London

Cahill J. (2010). *Flight Crew Task Performance and the Requirements for a New Tool Supporting the Pre-flight, Flight Planning and Briefing Task*. Unpublished doctoral dissertation, Trinity College Dublin, Dublin, Ireland.

Cahill J, McDonald N. and Losa G. (2008), '*Enhancing human performance and overall flight safety through improvements in airline system information flow'*. 28th Conference of the European Association for Aviation Psychology, Valencia, Spain, October 27-31, 2008.

Cahill J (2008), 'Envisioning Improved Work Practices and Associated Technology Requirements in the Context of the Broader Socio-technical System', In *Human Computer Interaction* (Open Access Book, In Press), I-tech Publishing, Vienna, Austria.

Carroll, J. M. (2000). *Making Use: Scenario-Based Design of Human-Computer Interactions*. MIT Press, Cambridge. MA.

Derrison, J. (2007). The SESAR Concept of Operations. Retrieved May 15, 2010 from, http://www.airtn.eu/documents/AirTN/AirTNForum1-6-Derisson.pdf

Eurocontrol (2009) Single European Sky. Retrieved 10 June 2009 from http://www.eurocontrol.int/ses/public/standard_page/sk_ses.html

European Commission (2001). European Transport Policy for 2010: Time to decide. Office for Official Publications of the European Communities: Luxembourg.

Group of Personalities (2001). *European Aeronautics – A Vision for 2020* - Advisory Council for Aeronautical Research in Europe (ACARE). Office for Official Publications of the European Communities: Luxembourg.

Johnson, W.W., Battiste, V., & Bochow, S.H. (1999). A cockpit display designed to enable limited flight deck separation responsibility. In *Proceedings of the SAE/AIAA 1999 World Aviation Conference*. Anaheim, CA.

McDonald, N., Corrigan, S., and Ward, M. (2002). *Well-intentioned people in dysfunctional systems*. Paper presented at the 5th Workshop on Human Error, Safety and Systems Development, Newcastle, Australia.

McDonald, N. (2006). Modelling the Human Role in Operational Systems. In, *Proceedings of the 17th European Safety and Reliability Conference on Safety and Reliability for Managing Risk. Estoril*, Portugal: Taylor and Francis.

McDonald, N. & Morrison, R. (2006). Modelling the Human Role in Operational Systems - Theory and Practice, 2nd Resilience Engineering Symposium, Antibes France, 8-10 November 2006.

Morrison, R. (2009). *Processing Mapping and the Knowledge Space Model:* Internal Report. Dublin: Aerospace Psychology Research Group, Trinity College, Dublin.

Muller, M. (2003). Participatory Design: The Third Space in HCI. In J Jacko & A. Sears (Eds), *The Human Computer Interaction Handbook: Evolving Technologies and Emerging Applications*. Lawrence Erlbaum Associates, Mahwah, NJ, 1051 – 1069.

SESAR Consortium (2008). SESAR Definition Phase. The Concept of Operations at a Glace. Brussels.

SESAR Consortium (2008). Definition Phase WP2.3 Performance Assessment – Document reference DLT-0706-231-00-05. Brussels.

SESAR Joint Undertaking. (2009). SESAR Brochure 02-04. Brussels.

Ward, M. (2005). *Contributions to Human Factors from Three Case Studies in Aviation Maintenance.* Unpublished doctoral dissertation, Trinity College Dublin, Dublin, Ireland.

DEVELOPING BEHAVIOUR-BASED SOLUTIONS THAT LAST EXAMPLES FROM INDUSTRY

Denis O'Hora[1], Nicole Gravina[2], Rachel Edwards[3] & Bruce Faulkner[4]

[1]*National University of Ireland, Galway,*
University Road,
Galway

[2]*Roosevelt University*
Chicago

[3]*Hollin Consulting*
Manchester

[4]*Three Simple Rules*
Manchester

Abstract

Organizational behaviour management (OBM) is an approach to enhancing the performance of organizations. Based originally on the theories of B. F. Skinner and other psychologists, OBM practitioners have developed tried and trusted solutions in the areas of behaviour-based safety, performance management and behavioural systems analysis. Core to an OBM approach is a focus on bringing about persistent and irreversible positive behaviour change. The current paper provides a brief conceptual background on OBM and details three examples of OBM-based solutions in a range of industries. Through these examples, we will demonstrate how organizational contexts inadvertently support and maintain substandard performance and how changing organizational contingencies and culture enables us to develop lasting behavioural solutions.

Introduction

Organizational behaviour management (OBM) is a scientific approach that grew out of the discipline of behaviour analysis. Based on the theories of B.F. Skinner and other psychologists, the applied sub disciplines of behaviour analysis focus on studying operant conditioning and respondent conditioning procedures that influence behaviour. The primary goal of applied researchers in this field is to predict and control socially important behaviours. Behaviour analysis typically uses an inductive approach to evaluate the impact of a planned change in the environment on behaviour over time. This means that the focus is on evaluating procedures rather than testing theories, resulting in empirically-based behaviour change intervention strategies.

The first applied behavioural interventions developed were related to programmed instruction and clinical applications, which were later followed by interventions in the workplace. OBM did not emerge as a sub discipline of behaviour analysis until the 1960's and in 1977 a journal aimed at publishing OBM research, Journal of Organizational Behaviour Management, was established. Since then, hundreds of OBM research studies have been published and a number of OBM consultancies have been established.

OBM researchers and practitioners primarily employ consequence-based interventions aimed at improving important work behaviours and results like productivity (e.g., Wilk & Redmon, 1999) and safety (e.g., Olson & Austin, 2001). Typically, behaviours and results are measured either through direct observation or their permanent products are evaluated after the event, rather than relying on indirect measures to assess performance (Daniels & Daniels, 2005). Some of the most common OBM interventions include training, feedback, recognition, and incentives (Nolan, Jarema, & Austin, 1999). An effort is made to permanently embed these procedures into the work environment, which often results in longer lasting improvements in performance (Sigurdsson & Austin, 2006). One of the most enduring OBM interventions published reported improved performance data across 11 years through an incentive-based intervention that improved safety of miners (Fox, Hopkins, & Anger, 1987)

In the current paper, we provide brief descriptions of three case studies that we have recently completed in a variety of contexts. The interventions we describe in each of these stem from theory and research in organizational behaviour management.

Case Study 1 – Coaching facilitated by data collection and feedback

Setting

In Case Study 1, a senior manager at a medium-sized organization request coaching. The manager requested assistance in reducing stress and managing workload. In conversation, the target of the coaching intervention was to increase the time that the manger spent on high value-added behaviours.

Data collection

Every 15 minutes, a small handheld device vibrated to prompt data entry. The client pressed a button to stop the vibration alarm, then entered data using a mobile phone application. The application presented a list of five activity categories. The client chose the activity category that she was currently engaged in, then pressed Send. Data was then sent via the mobile device to a remote database.

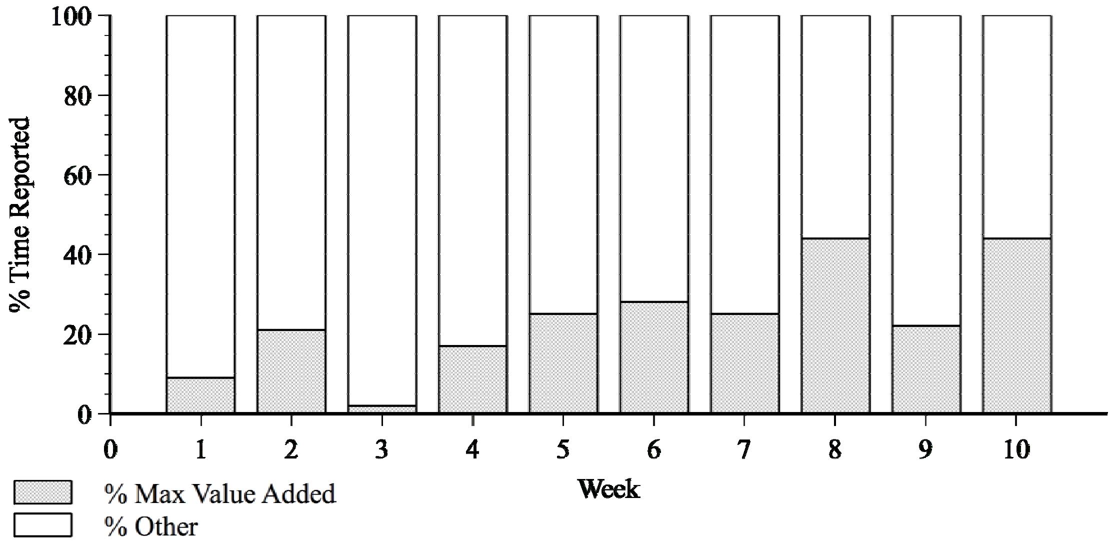


Figure 1. This figure depicts time spent on tasks categorised as Maximum Value-Added tasks as a percentage of time spent at work across the 10 weeks of the intervention

Results

In conversations with the client, certain categories were identified as High Value Added Activities (e.g., strategic planning). Figure 1 displays the client's data across the 10 weeks of data collection.

Case Study 2 – Public Posting of Employee Performance

Setting
A design manager at a small computer-focused enterprise requested assistance in reducing the time taken by employees to close out queries.

Data collection
Data were collected weekly.

Intervention
It was suggested that the design manager that post performance data and goals publicly. Such public posting works best in a supportive feedback-rich environment and the client organisation satisfied these criteria.

Results
Employee response rates increased rapidly (see Figure 2) when public posting was implemented. In addition to an increased speed of response to queries, a number of simple obstacles were found during the intervention period. By quickly removing these obstacles, the design manager facilitated more effective work practices, which helped to reinforce higher rates employee performance.

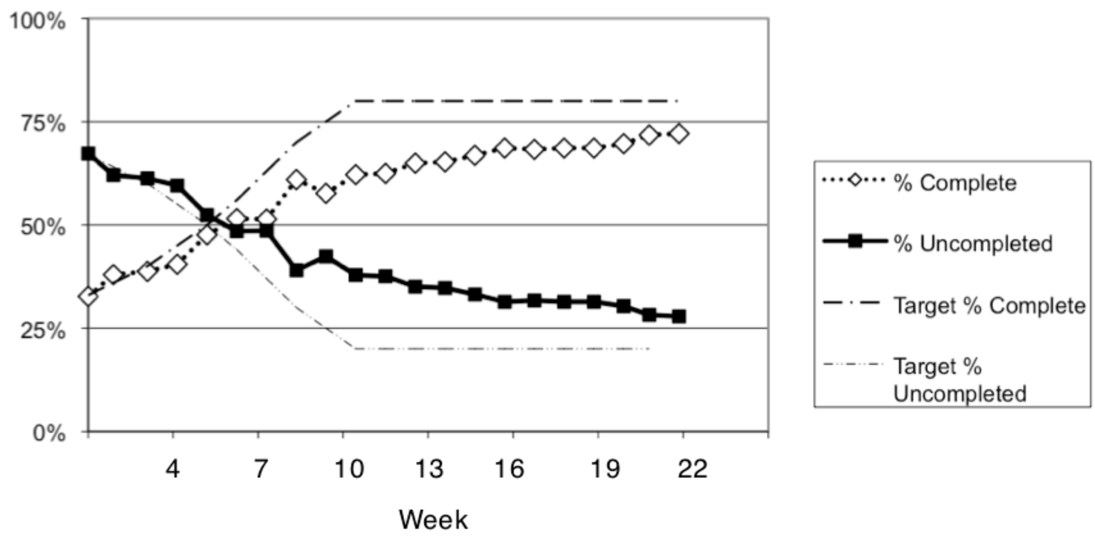


Figure 2. Online Response Target Performance across 22 weeks.

Case Study 3 – Employee-designed behaviour improvement plans

The foregoing case studies addressed relatively specific problems. In Case Study 1, the problem was self-identified and in Case Study 2, senior management recognized the problem. However, many companies operate sub-optimally, but senior management are unaware of specific problems that need to be solved. Senior management may wish to decrease injuries, decrease costs or increase profitability, but they may not have identified a specific issue or problem to be addressed. Box (1991) suggests that the vast majority of problems that occur in an organization can be solved relatively easily by those closest to the problem, specifically front-line employees.

Many OBM interventions focus on training front-line employees in the skills of observing, recording and analyzing behaviour in order to enhance employees' abilities both to identify problems on the ground and to solve them. Case Study 3 briefly summarises a training package that we provided and the results of that intervention.

Setting

Case Study 3 was conducted in a local authority in the United Kingdom. There was no explicit target behaviour that the client wished to change. In order to assess value to the client, the primary metrics were cost, quality (right first time), time spent, and reducing lost time.

Improvement intervention

231 employees at a range of grades in the organization were trained on an OBM course. The course consisted of half-day modules run every fortnight for around 3 months, with background reading, online coursework and ongoing assistance from a dedicated tutor. The product of the course was the implementation of an individual improvement project by each candidate in their local work environment. These projects were designed to produce savings and streamline processes at different levels of the organisation in order to meet the current challenges faced by the Council. Examples of these projects are outlined below.

Low Usage Vehicles

A Task Force Manager identified £450k of annual savings due to low or unused vehicles that were being leased by the Council. He established which managers were responsible for the leasing process and agreed new expectations with them about acceptable usage levels, and vehicles to be off-hired. Data was fed back and discussed at a monthly meeting where the managers accounted for their leased vehicles. After 3 months £150k had been saved.

Incorrectly Completed Energy Assessments

A Senior Quantity Surveyor was responsible for a team of surveyors who visited Council properties to undertake energy assessments. He found that many of the assessments the team submitted were missing information, so he filled them in himself; when he collected data, he found this accounted for 1 day of his working week. He set new expectations that the team would be re-visiting jobs to collect missing information, and reinforced progress by delivering the data on improvements. Within 2 weeks errors had fallen by 95%, saving £15k every quarter.

Repeat Domestic Bins Requests

A team manager identified a high level of repeat domestic bins requests from customers. He gathered data and found that there was a time-consuming 11-stage process for a new bin to be ordered; sometimes the customer would make a duplicate request if they hadn't received a replacement bin. The team manager found that neither the call centre nor the refuse supervisors were checking for duplicate orders in the system. He reset expectations to all and delivered weekly feedback on progress. Within 4 weeks duplicate requests had reduced to zero, saving £10k.

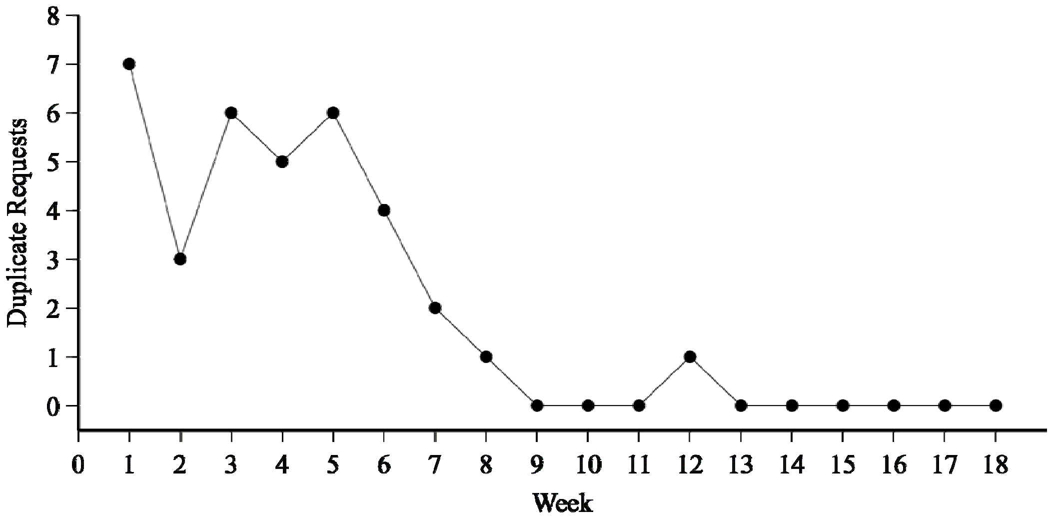


Figure 3. This figure shows the number of duplicate bin requests recorded across 18 weeks of the intervention

Acknowledgements

The authors are members of the BMT Federation, an international group of behavioural scientists and consultants. They would like to thank the other members of the BMT Federation for their support and insightful comments.

References

Box, G. E. P. 2006, *Improving almost anything: Ideas and essays.* Wiley-Interscience.

Daniels, A., and Daniels, J. 2005, *Performance management: Changing behaviour that drives organizational effectiveness.* (4th ed.) Atlanta, GA: Performance Management Publications, a division of Aubrey Daniels International, Inc.

Fox, D. K., Hopkins, B. L., and Anger, W. K. 1987, The long-term effects of a token economy on safety performance in open-pit mining. *Journal of Applied Behaviour Analysis*, 20(3), 215-224.

Nolan, T. V., Jarema, K. A., and Austin, J. 1999, An objective review of the Journal of Organizational Behaviour Management: 1987-1999. *Journal of Organizational Behaviour Management,* 19(3), 83-114.

Olson, R., and Austin, J. 2001, Behaviour-based safety and working alone: The effects of a self-monitoring package on the safe performance of bus operators. *Journal of Organizational Behaviour Management*, 21(3), 5-43.

Sigurdsson, O. S., and Austin, J. 2006, Institutionalization and response maintenance in organizational behaviour management. *Journal of Organizational Behavior Management*, 26(4), 41-77.

Wilk, L. A., and Redmon, W. K. 1997, The effects of feedback and goal setting on the productivity and satisfaction of university admissions staff. *Journal of Organizational Behavior Management*, 18(1), 45-68.

EVALUATING A COMPUTER SYSTEM THAT CONVERSES FLUENTLY WITH PEOPLE

Roddy Cowie and Gary McKeown

School of Psychology, Queen's University, Belfast

Abstract

The SAL system embodies a new kind of human-computer interaction, where a person and a computer carry out a fluent, emotionally coloured conversation. Because that kind of capability is new, evaluating systems that have it is a new challenge. This paper outlines techniques that have been developed to evaluate SAL interactions, and uses the case to highlight the range of variables that become relevant in dealing with systems of this order of complexity.

1 General background and aims

One of the natural long-term goals for human-machine interfaces is to develop systems that can engage a human being in a face-to-face conversation which is fluent, sustained, and emotionally coloured (Cowie et al 2001, Cowie & Schroeder 2005). The first artificial system that actually carries out such conversations has been developed by the FP7 project SEMAINE. It is called SAL, which is short for "Sensitive Artificial Listeners". Evaluating systems of that kind presents new challenges, and the subject of this paper is the evaluation of the SAL system.

The system itself has been described elsewhere (Schroeder 2010), but a brief outline is given here. Physically, the SAL user faces a teleprompter screen. He/she sees the face of a computer-generated agent on the screen, and hears its voice through speakers beside it. Behind the teleprompter screen are TV cameras, which feed the computer. From that and microphones, the computer extracts information about the user's emotions and nonverbal behaviour. Those drive the behaviour of the computer-generated agent, so that its facial expression, nods, backchannels, etc are attuned to the user's behaviour. The verbal component is a highly simplified device that allows this rich nonverbal behaviour to take place. SAL agents are defined by 'scripts' of stock utterances whose main function is to encourage or provoke the user to talk. They are modelled loosely on the conduct of a chat show host, who reveals very little, but encourages the other party to talk in an emotionally expressive way. The system simulates four 'characters', each with its own emotional style and agenda. Spike is bad-tempered, and his script tries to draw the user into a similarly bad-tempered view of the world; Obadiah is gloomy, and his script tries to draw the user into a similarly gloomy view of the world; Poppy is bright and cheerful, and her script tries to draw the user into a similarly chirpy view of the world; Prudence is sensible, and her script tries to draw the user into a similarly matter-of fact view of the world.

Building the system was a major effort of integrating the technology and the various human sciences that deal with face-to face interaction. The details are not relevant here. In brief, though, it allows a user who enters into the spirit of the system to interact in a way that is strikingly unlike traditional human-machine interaction – to the extent that users are often reluctant to believe that the system is truly automatic.

Central to understanding the system is the fact that it does not attempt to recognise very much of the verbal content of the user's speech: because it is fluent and emotionally coloured, that is beyond current technology. It is not necessary because the agent scripts are written specifically so that a suitable next utterance can be selected on the basis of the state of the conversation (e.g. ‘how was your day?’, ‘tell me more’, or ‘go for it’) or the user’s response (e.g. ‘I suppose you really think you’re something’, or ‘I love hearing all this happiness’). That makes the situation very unlike standard human-computer interaction, where transmitting information encoded in symbols (verbal or otherwise) is the object of the exercise, and evaluation focuses on the efficiency with which it is transmitted.

Evaluation of systems like SAL involves several different levels. As a first approximation, lower level issues can be separated out and addressed in comparatively straightforward ways — for instance, by measuring how often emotion is identified correctly from the user’s voice, or from his/her face, or from the combination. This report does not comment on those issues. Its focus is on the less concrete, but at least equally important high-level problem, which is to evaluate the system as a whole.

It is widely recognised that high-level evaluation of emotion-oriented systems presents particular challenges (Westerman et al 2006). SAL is one of the most purely affective systems in existence (if not the most). Hence evaluating the system is not simply about applying standard methods – it is about developing ways of dealing with various problems that have no agreed solution. Part of the challenge is that there was no way to know in advance what it would be like to hold conversations with an autonomous affective system. As a result, a key part of the evaluation process was identifying relevant variables – dependent variables that expressed the character of the experience, and independent variables that affected it.

This paper summarises the main points of the process. In order to do that, it covers early experiments which uncovered important problems as well as the final studies which confirm in a broad sense that SAL’s emotional competence enhances users’ experience.

2 Bases for selecting measures

There is a substantial literature on usability, which offers well-defined resources. However, it is underpinned by the conception of usability stated explicitly in ISO 9241, which defines it as the “extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.” (ISO 1998, p. 2]. Effectiveness and efficiency are not the issue in a system like SAL. Satisfaction has an affective component, but even it is defined in functional terms, as lack of discomfort, and a positive attitude towards the system, while performing the goals. Clearly that does not cover everything that people might look for in an affective system. As Edwardson put it, “We don’t ski to be satisfied, we want exhilaration” (1998, p. 2).

More richly affective measurement systems have been developed gradually, but they too often have specific goals in mind. Westerman et al. identified four main areas where measurement has developed: computer anxiety; trust and loyalty; frustration; and ’flow, fun, and playfulness’. The first two appear not to be relevant for SAL. Frustration clearly is, and one might assume it was simply undesirable. However, that is not necessarily so. There is a kind of frustration that is a mark of human engagement. If we treat them as people, then it is right and proper that we should be frustrated by Obadiah’s relentless pessimism or Poppy’s relentless brightness. Engagement is also a key issue in the last area. If Spike is convincing, an encounter with him is neither fun nor playful. However, it does create the characteristic ’flow’ feeling of being engrossed in a task, to the exclusion of distractions (see Csikszentmihalyi & Csikszentmihalyi, 1975).

Highlighting engagement points to another cognate area. One of the issues is clearly whether users feel that they are engaged in a conversation with a coherent personality, and respond realistically, despite knowing intellectually that the other party is not real. These are closely related to the issues that have been highlighted in research on presence (e.g. Sanchez-Vives and Slater 2005); and, of course, they are also related to the Turing Test (Schroeder & McKeown 2010).

A final point that recurs in the literature is scepticism about verbal measures (Isbister et al 2006). There are two main problems associated with them. One is that they disrupt the interaction that they are supposed to be describing. The other is that verbalising memories of emotion-related experiences involves quite radical transformations (Pennebaker & Chung, 2007), so that the reports systematically misrepresent the experiences.

On that basis, we identified six broad types of measure that it seemed appropriate to consider.

Competence Traditional evaluations focus on system competence. In the case of SAL, the scope for that kind of measure is limited. However, there is one obvious competence to consider, that is, conversational competence – whether the avatar appears to say appropriate things at the right time.

Sense of flow From the literature, probably the most promising measure is the sense of 'flow', which reflects a judgment that people make easily and reliably, even though its meaning may be hard to articulate.

Engagement This is related to flow, but there is a difference which is conceptually important. Engagement is specifically a state of the user, whereas flow is a characteristic of the activity – which is this case is jointly constructed by the user and the system.

Simple affect There are ways of measuring like and dislike which have the potential to be very misleading in this context – asking 'do you like Spike?' misses the point. However, measures at a more general level may well make sense. It seems plausible that people may be able to judge whether they like interacting with a system irrespective of whether they like the characters.

Non-verbal measures The literature does not suggest any obvious non-verbal measures, but logically it seems natural to ask for a response when the interaction is going badly. That turns an apparent problem to advantage. Giving the signal is a secondary task. Secondary tasks are routinely used as measures of engagement, because people who are immersed in the primary task tend to forget them. Hence it would be inappropriate to ask for a response when the interaction was going well, because the user would probably not be attending to the task when the response was appropriate. However, if the task is to make a secondary response when the interaction is going badly, then the times when attention is likely to be available are precisely those when it is needed.

Third party behavioural People who watch an interaction have a strong sense of whether the parties are engaged – reflecting the fact that there are rich behavioural indicators of engagement. Asking onlookers to record their sense of the user's engagement is a natural complement to the other techniques.

Another possibility in principle is to consider Turing-like tests, where the question is whether users can tell whether the system is autonomous or at least partly controlled by a human operator. That avenue was not pursued at this stage, for two main reasons. One is that it would become technically very elaborate, involving arrangements that would allow a human operator to control various aspects of the system's behaviour. The other, deeper reason is that the question is fundamentally a distraction. It is not the aim of the project to create avatars that will deceive people into thinking that they are human. To put resources into passing Turing-like tests would be to let the undoubted fascination of the test take priority over the project's natural goals.

On the basis of these considerations, we developed and explored a collection of techniques that seem to be suited to this particular evaluation task. They were developed through a series of pilot experiments.

3 Evaluations using Semi-automatic SAL

Semi-automatic SAL was a prototype for the eventual SAL system, where a SAL character's script consisted of a pre-recorded set of utterances, spoken by actor chosen to suit that character. A human operator chose which of the utterances the system would use at any given point in the conversation, using a moderately complex interface which allowed the operator to see a range of options that might be relevant, given the user's current state, and to choose by clicking a mouse which to use and when. The system's output is purely audio. Visually, the user saw a schematic face on the teleprompt, attached to a display that varied with the sound: its function is simply to provide a natural focus of attention.

The point of evaluating Semi-automatic SAL was to provide benchmarks against which the automatic system could be compared. The intention was to obtain different levels of performance by manipulating feedback to the human operator, so that ratings of the automatic SAL system could be located relative to ratings of various semi-automatic systems: at one extreme, a semi-automatic system controlled by a human operator with full feedback; and at the other, a system controlled by a human operator trying to conduct a conversation with virtually no information about what the operator was saying.

3.1 Method

Two sub-experiments were carried out. The first used relatively complex manipulations of feedback to the operator. In the full feedback condition, the operator could both see the user (via the cameras behind the teleprompter screen) and hear him/her (via loudspeakers). In the mid feedback condition, the visual channel remained, but the auditory channel was filtered by cutting out frequencies between 350Hz and 4000Hz: this gives a reasonable impression of prosody, but very few words can be made out. In the low feedback condition, the video channel was removed, leaving only the filtered audio. Interactions with the aggressive character, Spike, always used full feedback. Each participant also interacted with the other three characters, each with a different level of feedback. The different levels of feedback were balanced across the three characters.

In the second experiment, feedback was reduced by eliminating sound altogether, but leaving vision. Specifically, for two out of the four characters that each participant spoke to, the sound was switched off so that the operator could not hear what they were saying, and thus rely on this to select an appropriate response. Each conversation with the four characters lasted 20-30 minutes.

The basic evaluation used for both consisted of verbal reports by users. It is a known problem that asking for verbal reports during an interaction is likely to disrupt it; and on the other hand, reports given afterwards are likely to rationalise it. The solution adopted here was designed to minimise disruption by letting users respond from within the scenario. Immediately following each interaction, a different character (deliberately neutral) stepped in and asked three questions about the interaction that had just finished:
a) How naturally do you feel the conversation flowed?
b) How often did you feel the avatar said things completely out of place?
c) How much did you feel that you were involved in the conversation?
Answers to a) and c) were on a scale from 0 (worst) to 10 (best). Answers to b) were in terms of five categories, never (scored as 5); once or twice; a few times; quite often; most of the

time (scored as 1). Question b) was placed between the others because since it is in a different format, it reduces the tendency to answer the others in the same way.
In the second experiment, a non-verbal concurrent task was also used. Users were given a button to hold during the conversation, and were asked to press it whenever they felt that the simulation was not working well. It was christened a 'yuk button', and the name has stuck.

3.2 Results

Table 1 summarises the results of the first experiment. The key result is that level of feedback had no robust effect on ratings. Ratings on all the scales remained relatively close, and well within the upper half of the range, even when feedback to the operator was only filtered speech. This confirms an assumption that is critical to the project, which is that a SAL conversation can in principle be carried out without understanding the linguistic content of the user's utterances. Hence trying to design a system that can carry out a SAL-type dialogue without understanding the users' words is not asking the impossible.

Table 2 shows the broad pattern of results in the second experiment. The main outcome is to reinforce the findings of experiment 1. Average scores on all four measures – flow, correctness, engagement, and concurrent task (the 'yuk button') were close for full and degraded systems. It is clear that the scripts developed for SAL allow an operator to carry out an interaction that is reasonably acceptable to a user with a small amount of feedback, and specifically without understanding the verbal content of the user's speech.

Table 1: Results in the first experiment

	Average of Q1	Average of Q2	Average of Q3
Low			
Ob	6.25	3.50	7.25
Poppy	6.25	3.75	7.50
Prudence	8.00	4.75	9.00
Mid			
Ob	7.00	4.00	8.50
Poppy	6.00	4.25	7.00
Prudence	8.25	4.00	8.50
Full			
Ob	6.00	3.75	6.50
Poppy	7.25	3.75	9.00
Prudence	6.25	4.25	8.25
Spike	6.25	3.92	7.67

Table 2: Results in the second experiment

Measure	speaker	Mean/character	
		Degraded	Full
flow	Poppy	5.33	5.83
	Spike	3.6	6.25
	Obadiah	5	4
	Prudence	5.6	3.5
	All	4.885	4.89
correctness	Poppy	3.33	4.33
	Spike	2.40	3.75
	Obadiah	3.60	4.00
	Prudence	4.00	3.00
	All	2.67	2.23
engagement	Poppy	4.67	6.17
	Spike	3.8	6.25
	Obadiah	5.6	4.75
	Prudence	6.4	3.5
	All	5.12	5.17
Yuk button	Poppy	0	1
	Spike	1.6	0.5
	Obadiah	1.6	1
	Prudence	0.8	1.75
	All	1	1.06

Correlations were used to explore the relationships among the different measures. As might be expected, questions 1 and 3 were strongly correlated ($r>0.8$ for all four characters). Other relationships between indicators seem to be character-dependent. For two characters, Spike and Prudence, concurrent indicators and all three verbal measures seem to intercorrelate in a single global evaluation. For the others, the verbal probes of flow and involvement hang together, but the probe of competence and the concurrent task seem to be unrelated to them or to each other. It is important for work in this area to register that impressions may covary with one character but not with another: in other words, measurement regimes need to be wary of assuming that 'one size fits all'.

4 Evaluations using Automatic SAL

Evaluation of the automatic system involved three experiments. Each had partly separate aims, but taking them together produced an overall pattern, and that pattern is the focus of the description. In the nature of the work, each experiment generated a large amount of data, and a great many points of interest. The aim here is not to describe them exhaustively, but to convey the core points that each study adds to our understanding of the interactions with SAL-type agents and the task of evaluating them.

The basic design was that each user interacted with two versions of the system – the most competent version available, and a degraded version that lacked particular skills. The intended strategy was to study two types of degraded system, one in which the system's emotion detection skills had been disabled, and one in which its expressive abilities had been disabled. In practice, technological problems prevented the design from being fully realised. Carrying out the first experiment revealed 'bugs' in what was intended to be the full system. To a lesser extent, that was also true for the second experiment. It also highlighted the fact that some key expressive capabilities had not been disabled.

These complications are not trivial. They reflect the fact that with systems as complex as SAL, satisfactory evaluation depends on very extensive and complex interactions between the system developers and the system evaluators. In effect, evaluation has a dual role, to direct development as well as to measure achievement (Kaye eat al 2011), and the roles are not wholly separate.

4.1 Method

As in experiments 1 and 2, the core evaluation questions were kept within the scenario: they were asked by a distinct (non SAL) avatar between sessions with individual SAL characters. A preparatory briefing was also scripted and delivered at the start of each session by the same avatar.

In experiments 4 and 5, two additional measures were introduced. Because users experienced two versions of the system, it was possible to ask them at the end of the session which versions they had preferred. Ratings made by a third party were also introduced. That was done using technique used elsewhere in SEMAINE, which we have called tracing. A tracer uses a mouse to record some property the user's of the user's apparent state – in this case, apparent level of engagement – from moment to moment, on a continuous scale. All of these ratings were made by a single experienced tracer, who observed the interactions as they took place (audiovisually, but in a separate room). Users had a 'yuk button', but the interface with the other software was not reliable, and results from it are not reported.

The design counterbalanced which system (full or degraded) came first, and randomised the order in which a system presented the four characters. In total 90 participants were tested. 14 of these were in piloting and preparation stages. 16 participants took part in Experiment 3, 11 female and 5 male; 30 participants in Experiment 4, 19 female and 11 male; and 30 participants in Experiment 5, 24 female and 6 males.

4.2 Results

It is useful to start with a simple answer to a simple question, which is whether the SAL system's capacities materially affect user experience. The final experiment (with the most satisfactory versions of the full and degraded systems) indicates clearly that its expressive abilities do. At the most basic, simple affect measures showed a clear difference. 29 out of 30 users expressed a preference for one of the two systems, and 24 of them preferred the full system ($p<0.002$ on a sign test). On all of the more specific measures, analyses of variance show that the full system outperforms the inexpressive control. The main effect of system is significant for mean behavioural engagement ($F_{1,28}=14.7$, $p=0.001$); appropriateness of avatar contributions ($F_{1,28}= 5.7$, $p=0.025$); and flow and felt engagement ($F_{1,28} =12.0$, p=0.002: these were considered in a single analysis because of the close relationship between the underlying evaluations).

Table 3: Interactions that are predominantly in the upper 1/nth of the behavioural engagement scale

Region within which most of the interaction lies	% of interactions that meet the criterion, with:	
	the full automatic system	the inexpressive control system
Upper half (closer to "compelling sense of engagement" than to "absolutely no sense of engagement")	93%	86%
Upper 1/3 ("quite engaged" or better)	34%	6%
Upper 1/4	14%	0%
Upper 1/5	3%	0%

The behavioural engagement ratings provide a particularly revealing way of conveying what interacting with the system is like, because the scale has qualitative descriptors attached. It runs from "absolutely no sense of engagement" to "compelling sense of engagement". The boundary between bottom and middle third is marked "weakly engaged", and the boundary between middle and upper third is marked "quite engaged". Given that structure, a revealing test is to ask what proportion of interactions are rated the upper $1/n^{th}$ of the scale for at least half of their duration. The figures are shown in table 3. They capture what seems a reasonable assessment: the expressive abilities of the full system make it possible to sustain a quality of interaction that is unlikely to be achieved otherwise.

These statistics give a partial sense of a complex picture. Table 4 provides a broader summary of responses on the core measures, across all the experiments. The colour code is designed to bring out key patterns. In each row, the highest rating is coloured green; the next highest yellow; the lowest red; and the next lowest orange.

It stands out immediately that the worst ratings are concentrated in two columns, the first and the second last. The low ratings in the first column are informative because the operator is a human, meaning that the higher ratings in the later columns show the automatic system outperforming a human operator – albeit one whose feedback from the user involves only vision. The ratings in the second last come from the final degraded system, in which expressive functions have been disabled. The low ratings there confirm that the final SAL system's expressive capacity is important for its reception by human users.

Two particular cases cut across these general trends. The best ratings for Poppy come from experiment 2, where a human being with full feedback is choosing what the character says. The most striking point is that that only happens with one character. That is a considerable endorsement of the system. However, it is also interesting, and important, to ask why it happens with that particular character. It may be that a negative character can be convincing without referring directly to anything that the user is saying, but a positive character cannot.

A similar issue arises with Prudence. Her best ratings come in degraded systems, with experiment 3 best overall, and experiment 2 (with semiautomatic SAL) following. The natural explanation for the high ratings on experiment 3 is that Prudence is most acceptable when her constant reasonableness is broken by moments of irrationality – in particular, by statements that are quite at odds with the user's actual feelings. There is support for that in the very low ratings that she receives when a human operator with full feedback is choosing her utterances (in experiment 2). It seems that the more Prudence acts in character, the less people like her. That has important practical implications, because Prudence-like characters – projecting competence and rationality – are the sort that it is natural to imagine in applications.

Table 4 Overview of results with core measures

Measure	speaker	Semiautomatic 2		Experiment 3		Experiment 4		Experiment 5	
		Degrad	Full	Degrad	Full	Degrad	Full	Degrad	Full
flow	Poppy	5.33	5.83	4.63	4.57	5.12	5	3.93	5
	Spike	3.6	6.25	5.47	5.17	4.65	5.75	4.17	4.57
	Obadiah	5	4	5.77	4.77	4.77	5.98	4.75	6.05
	Prudence	5.6	3.5	6.2	4.5	4.38	4.33	4.43	5.07
	Overall	4.88	4.90	5.52	4.75	4.73	5.27	4.32	5.17
correctness	Poppy	3.33	4.33	3.20	3.80	3.73	3.82	3.40	3.42
	Spike	2.40	3.75	3.40	3.80	3.42	3.90	3.65	3.90
	Obadiah	3.60	4.00	3.73	3.73	3.75	3.88	3.62	4.22
	Prudence	4.00	3.00	3.79	4.07	3.78	3.66	3.37	3.70
	Overall	3.33	3.77	3.53	3.85	3.67	3.82	3.51	3.81
engagement	Poppy	4.67	6.17	5.47	5.86	5.8	5.82	5.22	6.15
	Spike	3.8	6.25	6.03	7.07	5.77	6.4	5.77	6.4
	Obadiah	5.6	4.75	6.73	6	5.95	6.35	5.42	7.08
	Prudence	6.4	3.5	6.8	6.27	5.23	5.52	5.5	6.08
	Overall	5.12	5.17	6.26	6.30	5.69	6.02	5.48	6.43

A more complex kind of character effect appears in responses to Spike. These appeared to be much more negative when he was the first SAL character encountered. Table 5 shows the effect in terms of flow ratings. By chance, Spike happened to be selected relatively often in first position in the last experiment, and that may well play a part in his relatively low overall ratings in that experiment.

Table 5 The 'Spike First' effect on flow

Experiment	Spike First (Mean)	Spike Other Position (Mean)	Spike First Frequency	N
3	1.75	5.57	2	15
4	3.17	5.43	6	30
5	1.6	4.92	10	30
Total	2.14	5.27	-	75

Those particular points underscore a general issues which is extremely important for the development of affective avatars, which is that avatar personality plays a major part in people's likes and dislikes. It is not simply a matter of which personality people like. Competences that make one personality more acceptable make another less so, and a personality that is difficult to deal with initially may be well received by people who are used to the system. These points can emerge because SEMAINE included avatars with contrasting personalities.

A parallel point that the techniques bring out has implications for any kind of evaluation. The user has to be recognised as a variable in evaluations of this kind of system. That is true at two levels. Informally it seems clear that enduring traits of the user have a substantial effect on response. It seems likely that some people will not engage in this kind of situation however socially skilled the system. Formal measures of personality were used to explore that issue, but the results were mixed, and will not be discussed. Transient states also have an effect. On one hand, people adjust to the unusual kind of interaction. On the other, first interactions in particular can dispose people positively or negatively towards the system. One would expect formal evaluations to balance for effects like these, but they also need to be taken into account in design-oriented studies. It would not be reasonable to assume that the reactions of a particular individual, perhaps with extensive experience of the system, could stand for the population at large.

These issues are food for a great deal of thought. However, it is important not to lose sight of the most basic finding of the evaluation. It is simply that the SEMAINE system clearly can engage people in sustained, emotionally coloured conversations. The system is deeply limited, but it demonstrates that emotion-rich, spoken conversations with avatars are a real possibility rather than science fiction.

5 Conclusion

The SAL system marks a new departure in the kind of interaction that can occur between human beings and machines. Correspondingly, evaluating it raised challenges which are different from those that HCI has well-established methods of addressing. The work reported here brings together a variety of techniques and concepts that seem suited to the task.

Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 211486 (SEMAINE).

References

R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human computer interaction," IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32–80, 2001.

R. Cowie and M. Schr¨oder, "Piecing together the emotion jigsaw," Machine Learning For Multimodal Interaction, pp. 305–317, 2005.

M. Csikszentmihalyi and I. S. Csikszentmihalyi, Beyond boredom and anxiety. San Francisco, USA: Jossey-Bass, 1975.

M. Edwardson, "Measuring consumer emotions in service encounters: an exploratory analysis," Australasian Journal of Market Research, vol. 6, no. 2, p. 34–48, 1998.

K. Isbister, K. Hook, M. Sharp, and J. Laaksolahti, "The sensual evaluation instrument: developing an affective evaluation tool," in Proceedings of the SIGCHI conference on Human Factors in computing systems. Montreal, Quebec, Canada: ACM, 2006, pp. 1163–1172. [Online]. Available:
http://portal.acm.org/citation.cfm?id=1124772.1124946

ISO 1998 I. I. O. for Standardization, "Ergonomic requirements for office work with visual display terminals (VDTs) - part 11: Guidance on usability," International Standards Organisation, ISO Standard ISO 9241-11, 1998.

Pennebaker, J. W., & Chung, C. K. (2007) Expressive Writing, emotional upheavals, and Health In H. S. Friedman & R.C Silver (ed) (Ed.), Foundations of health psychology. New York, NY: Oxford University Press pp. 263-284

M. V. Sanchez-Vives and M. Slater, “From presence to consciousness through virtual reality,” Nature Reviews Neuroscience, vol. 6, no. 4, pp. 332–339, 2005. [Online]. Available: http://dx.doi.org/10.1038/nrn1651

M. Schröder, “The SEMAINE API: Towards a standards-based framework for building emotion-oriented systems”, Advances in Human-Computer Interaction Volume 2010 (2010), Article ID 319406 doi:10.1155/2010/319406

M. Schroeder & G. McKeown Considering social and emotional artificial intelligence. Proc AISB symposium “Towards a comprehensive Turing Test”. Leicester, 2010.

S. Westerman, P. Gardner, and E. Sutherland, “Usability testing Emotion-Oriented computing systems: Psychometric assessment,” HUMAINE deliverable D9f, 2006. [Online]. Available: http://emotion-research.net/projects/humaine/deliverables/D9f%20Psychometrics%20-%20Final%20-%20with%20updated%20references.pdf

MOTIVATION IS FOR PEOPLE WHO AREN'T ALREADY HAVING FUN

Raymy Kate O' Flynn and Thomas Waldmann

Design and Manufacturing Technology Department
University of Limerick,
Castletroy
Limerick

Abstract

The Iowa Gambling Task (IGT), is a task that simulates real world decision making allowing for investigation into advantageous and disadvantageous behaviour. Five out of nine participants were told that the best performer on the IGT would receive a monetary reward of €30 while four others were not. The averaged data seemed to show that an extrinsic reward had a detrimental overall effect on performance. However analysis of individual data showed that most performed better in the task when an extrinsic reward was offered. To foster motivation, individuals need to internalize three basic concepts. To be intrinsically motivated and one must have a sense of autonomy, mastery, and purpose.

Introduction

In the IOWA Gambling Task (IGT) participants are confronted with four decks of cards, two of which are "advantageous", leading to moderate gains and losses, and a positive balance. The other two decks are "disadvantageous", with higher gains and losses and a negative overall balance. There are usually 100 trials per participant. Using the ubiquitous formula for establishing advantageous behaviour (Bechara, Damasio Tranel and Damasio, 1997) we calculated participants' advantageous behaviour per block of the game. The formula allowed us to calculate the average of advantageous deck selections in the block and then subtract the disadvantageous deck selections in the block giving us either a positive or negative number. These scores and participants' preference we observed in more than 100 participants (O'Flynn and Waldmann 2010) corresponded to results reported in the literature for adults without brain injury (Bechara and Damasio 2004), Bechara, Damasio and Damasio (2000), Demareea, Burns and DeDonnoa (2009), Fry, Greenop, Turnbull, and Bowman, (2009), Hooper, Luciana, Conklin, and Yarger, (2004).

Unfortunately the behaviour during the IGT is very hard to interpret for participants and experimenter alike. Participants are confronted with four "one-armed bandit tasks" at once (Deck A,B,C and D), which makes the task highly complex. What happens to the participants depends on what they chose to do within the game. The traditional way of interpreting IGT data has been based on averaged responses over the nine "rounds" in the experiment. The scores for rounds are sometimes added together to give another overall score for "blocks".

Google Scholar lists about 15,900 studies involving the Iowa Gambling Task (IGT), with results presented in terms of rounds or blocks, or both. Invariably, participants' knowledge of the task was investigated by questionnaire after the testing session.

In a previous study (O'Flynn and Waldmann, 2010) we compared a number of students on various courses where we assumed that differences in performance could be explained by different levels of knowledge. However, we found that knowledge about contingencies did not predict behaviour. If knowledge does not predict behaviour on the IGT, what motivated people to do well on the task?

Method

Nine Participants ranged in age from 21- 28 with an average age of 22.8 (STDEV: 2.9). The participants were selected at random for one of two groups; Reward (5 participants;) and No Reward (4 participants). €30 was offered as a reward to the one participant in the Reward group who had the highest overall money balance at the end of the game. No mention of this prize was made to the No Reward group. All participants attended a Product Design and Technology (PDT) course at the University of Limerick.

The study used a computerised version of the IGT and computerised version of Maia and McCelland's (2004) Somatic Marker Questionnaire (SMQ). Facsimile rather than real money was the intended reinforcement during the game. Participants were allowed to select cards, in any order, from any of the four decks (A, B, C, and D). As in the original IGT decks A and B were disadvantageous decks and decks C and D were advantageous. As in the original IGT losses were more frequent on selection of decks A and B, and the task was terminated after 100 selections. Maia and McCelland's (2004) SMQ on participants' knowledge was presented on the computer screen and three levels of knowledge were calculated according to the SMQ.

Based on Maia and McCelland's (2004) definition participants were tested for their behaviour (selecting from more advantageous decks C and D on average during a trial), and on three knowledge levels. As in Bechara et al. (1994) initially the 100 card selections were sub-divided into five blocks and each block was calculated by subtracting the number of good from bad card selections ((C + D) – (A + B)). A net score above zero implied that the participants were selecting cards advantageously, and if participants scored a net score below zero their selection was considered disadvantageous. Net advantageous decisions were then calculated across five blocks of trials Participants' knowledge throughout the task was recorded after the initial twenty and then after every ten selections.

Results

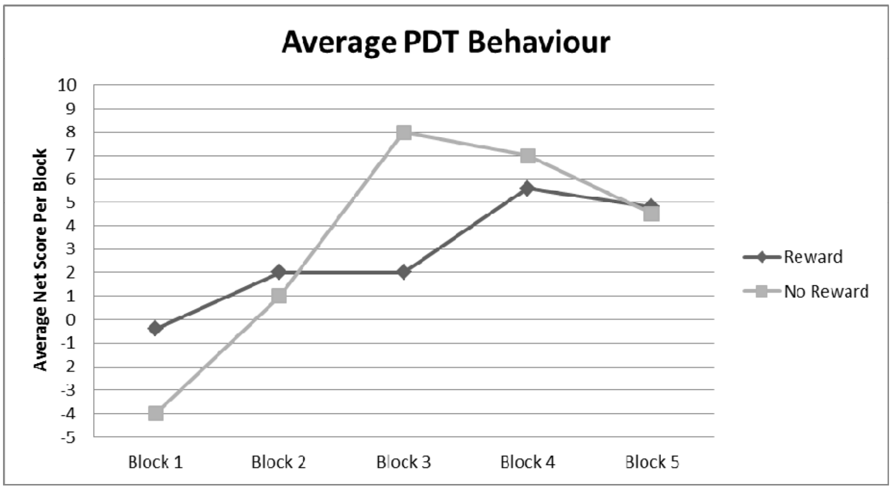


Figure 1 – Participants' advantageous behaviour averaged across 5 Blocks.

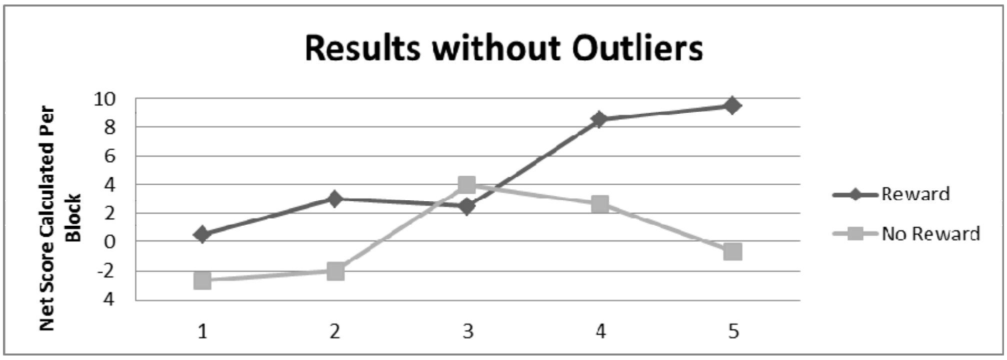


Figure 2- Participants' average results displayed with outliers excluded.

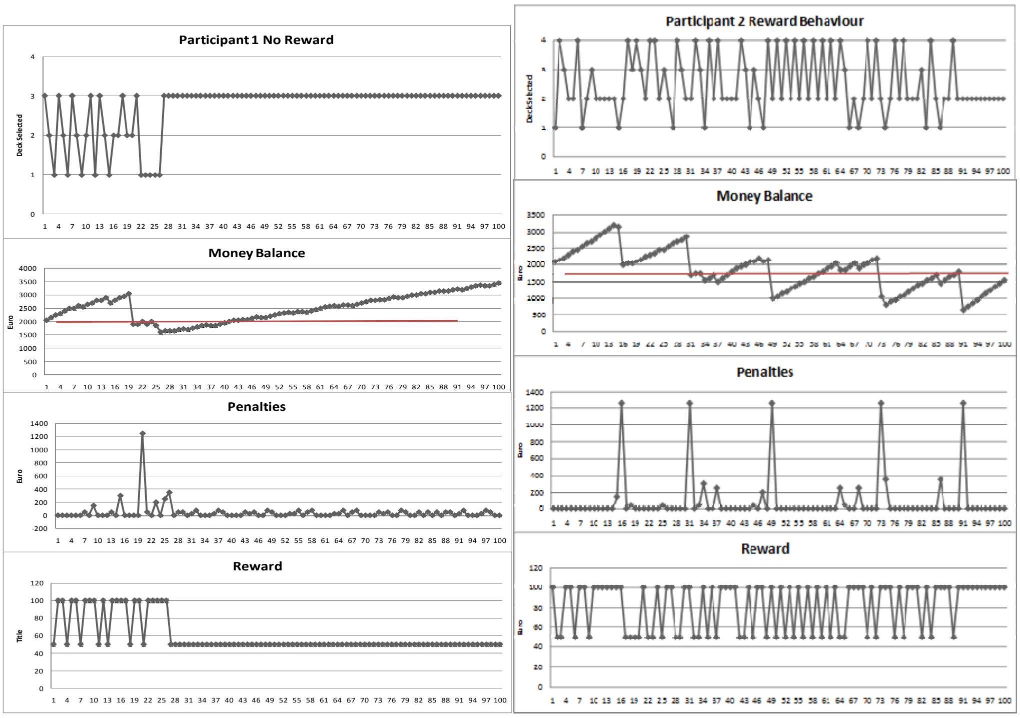


Figure 3- Participant NR1 and Participant R2. These were outliers within the game. Participants displayed exceptionally successful (NR1) or unsuccessful (R2) behaviour.

Discussion

Fernie (2007 p.88) stated that “reinforcer type does not have any effect on IGT performance”. Upon inspection of our data, we could have said the same based on Figure 1, however on the removal of outliers, individual data clearly show difference in advantageous decision making as a result of extrinsic motivation in the form of a €30 reward, as Figure 2 shows.

A monetary incentive improved performance for most people. There are generally two search strategies: some people stop searching when it seems to them that they have found a good enough strategy (see NR1, never bothering with deck D), others attempt to investigate every possibility. It is possible that the Reward group predominantly used a strategy of risk avoidance. Therefore those faced with gains (the Reward group) tend to be risk averse, while those confronting losses (the No Reward group) become much more risk seeking (McDermott, Fowler and Smirnov 2008). The “Hedonic Treadmill” is a good example of this (Brickman and Campbell 1971). People have to work harder to maintain the same level of happiness as experienced previously because human attention evolved to focus on change as opposed to equilibrium; sensitivity change that threatens to make things worse for individuals is therefore a priority for most.

In 128 studies reported by Pink (2009) rewards had negative effects on motivation. For example, for most employees the expectation that creativity will be especially rewarded causes the employee to define the task as requiring creativity, to become immersed in it, and to search for novel ways of carrying it out. Work by Deci (1971) suggests that once people have done enough to get paid they are less likely to keep working. In a meta analytic study conducted by Deci, Koestner and Ryan (1999) it was noted that engagement-contingent, completion-contingent, and performance-contingent rewards significantly undermined free-choice intrinsic motivation, as did all rewards, all tangible rewards, and all expected rewards. However, Nickerson (1985) suggests rewards can enhance creativity through increased intrinsic task interest but it must be dealt with sensitively. Reward for high performance increases perceived self-efficacy and perceived competence, both of which increase enjoyment of a task for its own sake. In contrast, the expectation that reward depends on conventional performance causes people to define the task as involving conventional performance, hampering creativity. Therefore it is important to frame the task in such a way as to provide emphasis on reward for high performance or for being creative in order to translate intrinsic interest into creative achievement.

References

Bechara, A., Damasio A. R., Damasio, H., and Anderson, S. W. 1994. "Insensitivity to future consequences following damage to prefrontal cortex." Cognition 50: 7-15.

Bechara, A., Damasio, H., Tranel, D., Damasio A. R., 1997. Deciding Advantageously Before Knowing the Advantageous Stragegy. Science 28 February: 275(5304), 1293-1295.

Bechara, A., Damasio, H., Damasio, A.,2000. "Emotion, Decision Making and the Orbitofrontal Cortex." Cereb. Cortex 103: 295-307.

Bechara, A. and A. R. Damasio 2004. "The Somatic Marker Hypothesis: A Neural Theory of Economic Decision." Elsevier: Games and Economic Behaviour 52: 336-372.

Brickman, P., Campbell, D., 1971. “Hedonic Relativism and Planning in Good Society” In M.H. Apley,ed., AdaptationLevel Theory: A Symposium, New York: Academic Press, pp 287-302.

Deci, E. L. (1971) Effects of externally mediated rewards on intrinsic motivation. Journal of Personality and Social Psychology, Vol 18(1), 105-115

Deci, E. L., Koestner R., Ryan R. M. (1999) a Meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychology Bulletin, Vol 125(6), 627-668.

Demareea, H. A., K. J. Burns, M. A. DeDonnoa 2009. "Intelligence, but not emotional intelligence, predicts Iowa Gambling Task performance." Elsevier Intelligence 382: 249-254.

Eisenberger, R. and Rhoades, L. (2001). Incremental effects of reward on creativity. Journal of Personality and Social Psychology, 81, 728-741.

Fernie, G. (2007). Factors affecting learning and decision making in the Iowa Gambling Task. PhD. Thesis: University of Nottingham, UK.

Fry, Jessica; Greenop, Kirston; Turnbull, Oliver and Bowman, Caroline (2009) The effect of education and gender on emotion-based decision-making. South African Journal of Psychology, Vol 39, Issue 1, Mar, pp. 122-132.

Hooper, C. J., M. Luciana, H. M. Conklin, R. S. Yarger 2004. Adolescents' Performance on the Iowa Gambling Task: Implications for the Development of Decision Making and Ventromedial Prefrontal Cortex. Developmental Psychology, American Psychological Association. 40: 1148-1158.

Maia, T. V. and J. L. McClelland (2004). "A Reexamination of the Evidence for the Somatic Marker Hypothesis: What Participants Really Know in the Iowa Gambling Task." Proceedings of the National Academy of Sciences of the United States of America 10145: 16075-16080.

McDermott R., Fowler J.H., Smirnov O. (2008): On the Evolutionary Origin of Prospect Theory Preferences. The Journal of Politics 70:335-350.

Nickerson, R. S. (1985). Enhancing creativity. In R. J. Sternberg (Ed), The nature of creativity(pp.11-38). Cambridge, UK: Cambridge University Press.

O’Flynn, R.K. and Waldmann, T 2010, 40th Anniversary Conference of the Psychological Society of Ireland , Wexford, Ireland.

Pink, D.H. (2009). Drive: The Surprising Truth about What Motivates Us. Riverhead Books: New York. Pp. 23- 28

AN ERGONOMICS TRAINING MODEL FOR PROCESS DESIGNERS

Aileen Jackson and Leonard O'Sullivan

Ergonomics Research Group
Enterprise Engineering Research Centre
University of Limerick
Limerick

Poster Abstract

Ergonomics has been gaining in ever increasing importance in many areas: from the design of products/services, to the prevention/reduction of safety and health problems, right down to human-machine interaction problems. Product, rather than Process Design, seems to be the focus of most training and models in existence though Process Design comes ever more to the fore when considering the design of production lines or in the more general arena of workplace design. This is due to the fact that the individual worker is the central element in the workplace.

Thus the challenge is to train students and professionals in adaptive, quick-response, process ergonomics design and evaluation, such that going forward the manufacturing process improves the quality, productivity and health of the worker.

ERGOMAN, a Leonardo project funded by the European Commission, has three main objectives. Firstly, a **Competence Needs Analysis** carried out in industry. This analysis aims to direct a Training Model toward the application of ergonomics design for production processes in a more effective way than today. The investigation is conducted through observation, interviews, questionnaires sent out to those working in ergonomics and also via focus group in different companies.

Secondly, a new **Competence Profile** for designers of ergonomic production processes and workplaces is established. Three competence profiles have been identified for the three targets groups of the project (students, professionals in big industry and process designers in industry).

Thirdly and finally, the creation of a suitable **Training Model** to train the different targets groups on the identified 'further training required' knowledge areas. The results of the project are to be used Europe-Wide because the professional profile has been designed to respond to industrial needs and standards throughout the European Union.

The exploitation plan for the Training Model involves proposals to develop e-learning modules in a **blended learning** framework using Softchalk software. It is proposed to host the blended learning version of the training model using Moodle, a shareware Learning Content Management System (LCMS) wish access to third level institutions who can access and use the material in their instruction.

Acknowledgements:
The ERGOMAN project "*New Designers for Production Processes Highly Ergonomic and Safe*" project number 503041-LLP-1-2009-1- IT-LEONARDO-LMP, is funded with the support of the Leonardo Da Vinci Programme of the European Union.