Measuring the Performance of the IT Function in the UK Health Service Using a Balanced Scorecard Approach

Maurice Atkinson
Southern Health and Social Services Board, Northern Ireland
mauricat@shssb.n-i.nhs.uk

Abstract: This paper explores how the Balanced Scorecard approach might be applied to measuring the performance of an IT department. Sample measures have been developed for each dimension of the scorecard for two key IT functions. A performance measurement record sheet has been developed to show how these measures would work in practice. The paper also outlines approaches to implementing, monitoring and reviewing these measures. Furthermore the benefits of such a performance management system and process have been identified.

Keywords: Information Technology, Balanced Scorecard, Performance Measurement.

1. Introduction

In an article written in 1991 Robert Eccles heralded a performance measurement revolution and predicted that "*within the next five years, every company will have to redesign how it measures its business performance".* This prediction has proved correct and "*the shift from treating financial figures as the foundation for performance measurement to treating them as one among a broader set of measures*" (Eccles, 1991, p131) has been embraced, to a greater or lesser extent, by both the private and public sectors. In this paper I will explore how "*a broader set of [performance] measures*" might be developed and implemented for an IT department within the context of the UK Health Service.

2. Context

The foundations of the current National Programme for IT (NPfIT) in the National Health Service (NHS) were originally set out in the Department of Health's strategy *Information for Health (*1998). This committed the NHS to:

- Lifelong electronic health records for every person in the country;
- Round-the-clock online access to patient records and information about best clinical practice, for all NHS clinicians;
- Genuinely seamless care for patients through GP's, hospitals and community services sharing information across the NHS information highway;
- Fast convenient public access to information and care through online information services and telemedicine; and
- The effective use of NHS resources by providing health planners and managers with the information they need.

Following the development of the Government's ten-year plan for the NHS (2000), a supporting document *Implementing the NHS Plan – Building the Information Core* was published in January 2001. This developed the vision set out in the NHS Plan for a "*service designed around the patient*". The delivery of a modernised health service was inextricably linked to the delivery of modern information technology as a key enabler of bringing about change in the way care is accessed and delivered. The Wanless Report (2002) included several key recommendations for IT in the NHS:

- A doubling of IT spending that is protected to ensure that it is not diverted for other purposes;
- Stringent national standards for data and IT set by the centre; and
- Better management of IT implementation, including a national programme.

The report underlined the fact that the NHS Plan's targets and modernisation objectives require a different, new approach to the development and delivery of IT:

> *"Without a major advance in the effective use of ICT, the Health Service will find it increasingly difficult to deliver the efficient, high quality service which the public will demand. This is a major priority which will have a crucial impact on the health service over future years."*

The NPfIT commenced in October 2002 and *Delivering 21st century IT support to the NHS* started the process of connecting the delivery of the NHS Plan and the modernisation of services to the information strategy. IT will now be designed and delivered around the needs of the patient and service users, not institutions; shifting from systems running

along institutional lines, dealing with only a portion of patient interactions, to whole health and social care community systems that track and record a whole user/patient journey. The key component of NPfIT is the building and availability of a 24/7 live patient record that all health professionals in whatever setting (hospital, primary care, community services) and patients can access.

Thus the importance of ICT in supporting the core business of an improved healthcare system and realising the targets and vision of the NHS Plan are beginning to be increasingly recognised. The priority which attaches to the contribution of ICT to the work of the NHS raises issues of how best to link ICT targets with business objectives and in turn determining how well the IT function is performing. One potential solution is the use of the Balanced Scorecard.

3. Performance Measurement

Neely (1998, p5-6) offers a definition of a performance management system as one which "enables informed decisions to be made and actions to be taken because it quantifies the efficiency and effectiveness of past actions through the acquisition, collation, sorting, analysis, interpretation and dissemination of appropriate data". In this context "effectiveness refers to the extent to which customer requirements are met, and efficiency is a measure of how economically the organisation's resources are utilised when providing a given level of customer satisfaction". The litmus test of a performance management system might thus be regarded as the use that is made of it to inform decisions and bring about improvements.

Performance frameworks fall into two broad categories: those which are designed for assessing business excellence e.g. the European Excellence Model (EEM), the Deming Prize of Japan, and the US Malcolm Baldrige National Quality Award; and those which are designed to help organisations develop performance management systems. The EEM is a diagnostic and self-evaluation tool whereas the focus of the Balanced Scorecard (BSC) is on providing a methodology to allow an organisation to turn its strategy into actual achievements.

The Balanced Scorecard (BSC) will be used in this paper as a conceptual and pragmatic model for the development of performance measures for an IT Department. The Balanced Scorecard (BSC) is a technique developed by Kaplan and Norton (1992) that helps organisational decision-makers to navigate the organisation towards success. It enables organisations to translate their mission and strategy into a comprehensive set of performance measures that provide the framework for a strategic measurement and management system.

The Balanced Scorecard measures organisational performance, with emphasis on financial objectives. However, it also includes the performance drivers of these financial objectives, and measures organisational performance across four balanced perspectives:

- Financial;
- Customer;
- Internal Business Processes; and
- Learning and Growth.

This is illustrated in Figure 1 below:

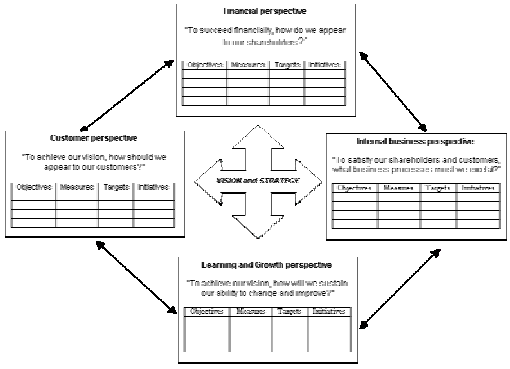

Figure 1: Kaplan and Norton's Balanced Scorecard (Kaplan and Norton (1996a, 1996b))

Kaplan and Norton (1992) argue that traditional financial measures are backward looking. They try to address this inadequacy by complementing past performance measures (financial measures) with drivers of future performance indicators (customers, suppliers, employees, processes, technologies and innovation). The fundamental concept of the BSC is to derive the objectives and measures from the overall corporate vision and strategy and to use four perspectives as a "balanced" framework to monitor and achieve these objectives.

The Balanced Scorecard, with its four perspectives, has been criticised (e.g. Neely, 1998) for excluding employees inside the organisation, suppliers and other external partners, competitors, and regulators. For example, the supplier perspective would be extremely important in the IT field where services may be outsourced or where a supplier acts as a partner within a system

procurement or application development project.

4. Identification of performance measures for IT

The BSC framework will be used to develop financial and non-financial measures for an IT department. Two areas of IT activity have been selected to illustrate how this approach might be applied in practice i.e.:

- Project Management; and
- IT Help Desk.

Critical success factors will be developed for each area together with performance measures within each of the 4 perspectives of the BSC.

The performance management process for IT which underpins the development of these performance measures involves linking IT objectives to the corporate strategy, following a BSC approach to identify target measures, implementing and monitoring these measures in order to improve decision making and bring about improvements in IT processes and performance.

4.1 Project management

A formal project management methodology such as PRINCE II (Projects in Controlled Environments) may be used to manage ICT projects. One of the key objectives of an IT department is to identify and prioritise projects in line with the ICT strategy and corporate plan, and to adopt and apply sound project management techniques for each project undertaken. Projects may relate to a wide range of areas e.g. internet/intranet development, system development and implementation, infrastructure development etc. etc.

Critical success factors (CSF's) for project management include:

- Experienced and skilled project managers are available;
- There is senior management sponsorship of projects;
- Stakeholders and IT staff share in the definition, implementation and management of projects;
- A project organisation is in place with documented roles and responsibilities;
- There is an understanding of the abilities and limitations of the organisation and the IT function in managing large, complex projects;
- All projects have a Project Initiation Document which includes project background and justification, Project Definition, Project Plan, Communication Plan, Project Quality Plan, Project Controls and Risk Log; and
- The transition from the implementation team to the operational team is a well-managed process.

It is critical that the effectiveness and efficiency of projects are monitored utilising key performance measures. The measures developed for project management are in 4 linked areas - Financial, Internal Business Process Learning & Growth, and Customer – and this is illustrated in Table 1 below:

Table 1: Performance measures for project management

BSC Perspective	Sample Performance Measures
Financial	Availability of accurate project schedule and budget information; Number of project milestones and budget reviews; Increased number of projects completed on time and on budget.
Internal Business Process	Increased number of projects delivered in accordance with a defined methodology; % of projects with post-project reviews; Decrease in systematic and common project problems; Improved timeliness of project management decisions.
Learning and Growth	Number of project management training days per project team member; Average number of year's experience of project managers.
Customer	Increased organisational satisfaction with project delivered services; % of stakeholder participation in projects.

The ultimate aim in this area is that a proven, full life-cycle project methodology is implemented and enforced, and is integrated into the culture of the entire organisation.

Kaplan and Norton (1996b, p30) argue that "*a strategy is a set of hypotheses about cause and effect. The measurement system should make the relationships (hypotheses) among objectives (and measures) in the various perspectives explicit so that they can be managed and validated. The chain of cause and effect should pervade all four perspectives".* The measures that have been identified for Project Management have been selected with this "*chain of cause and effect relationships*" in mind. Thus the measures act as indicators of the department's progress

towards meeting the overall objective and CSF's for the Project Management process and are causally linked to one another. For example, the increased use of a formal project management methodology will drive the move towards more project team members being fully trained in project management methodology which will lead to an increasing number of projects being delivered on time and to budget and a higher level of organisational satisfaction with project delivered services.

4.2 IT help desk

The IT Help Desk provides first-line support and advice to users of IT systems in an organisation with the aim of ensuring that any problem experienced by the user is appropriately resolved.

Critical success factors for the operation of the IT Help Desk include:

- Knowledgeable and customer-orientated support staff resolve problems in close co-operation with senior IT staff;
- All user enquiries are consistently and thoroughly registered by the Help Desk;
- User enquiries that cannot be resolved in a timely manner are appropriately escalated;
- The clearance of user enquiries is monitored;
- User questions are resolved in a timely manner;
- Those user enquiries that cannot be resolved in a timely manner are investigated and acted upon;
- Management monitors trends to identify root causes in a proactive manner and follows up with analysis and the development of solutions;
- Organisational policies and programmes are defined for training users in technology and security practices; and
- There is management awareness of support costs and these are charged back to the business.

The measures developed for the IT Help Desk are in 4 linked areas - Financial, Internal Business Process Learning & Growth, and Customer – and this is illustrated in Table 2 below:

Table 2: Performance measures for the it help desk

BSC Perspective	Sample Performance Measures
Financial	Cost per call.
Internal Business Process	Number of enquiries; Reduced average time to resolve problems; Reduced repetitive enquiries on solved problems; % of problems solved at first contact; Elapsed time per call; Number of escalations; Reduced trends in user enquiries requiring problem resolution.
Learning and Growth	Number of IT training programmes attended by staff.
Customer	Increased user satisfaction with the effectiveness and efficiency of the Help Desk; Increased user confidence in the services of the Help Desk.

The ultimate aim in this area is that the Help Desk function is established, well organised and takes on a customer service orientation, by being knowledgeable, customer focused and helpful.

As previously noted the performance measures for the IT Help Desk have been selected with Kaplan and Norton's (1996b, p30) "*chain of cause and effect relationships*" in mind. The measures act as indicators of the department's progress towards meeting the overall objective and CSF's for the IT Help Desk function and are causally linked to one another. For example, customer (and organisational) satisfaction is likely to increase if the time to resolve IT problems is minimised and the Help Desk service is delivered in a cost effective way. Furthermore, the need to effectively utilise the Help Desk resource will drive the provision of IT training programmes for staff in order to reduce the inappropriate use of the IT Help Desk. A more effective service could potentially be provided at higher cost, but judgements would need to be made as to the balance between the required level of service versus the resource available to provide this.

5. How the measures will work

In considering how the IT measures which have been identified will work in practice it is important to bear in mind the behaviour they will encourage and whether or not this behaviour is desirable. Neely *et al* (1996, 1997), Neely (1998), and Bourne (2000) have

developed a performance measurement record sheet which forces a series of questions to be answered in order to ensure that the measure is of practical value. There are 10 dimensions within this framework:

- Title of the measure;
- The purpose of the measure;
- What corporate objective/s the measure relates to;
- What performance target should be set;
- How is the performance measure to be calculated;
- Frequency of measurement and review;
- Identification of who is responsible for measuring performance;
- Source of the data;
- Allocation of responsibility for taking action on the measure; and
- Specification in outline of the types of action that can be taken to improve performance in
- this area.

The performance measurement record sheet provides a formal framework which explicitly links measures to objectives, further defines the measures and how they will be evaluated, assigns responsibilities, and ensures that performance improvement is integrated as part of the overall process.

In Table 3 (at the end of the paper) a Performance Measurement Record Sheet has been completed for the six of the IT measures previously identified as a way of demonstrating how they would work in practice:

6. Implementation

Olve and Sjostrand (2002, p106) emphasise the importance of what they call "*the virtuous circle formed by strategy, control, measurement, learning, and back to strategy".* If, as part of the implementation of a performance measurement system, this continuous process is maintained then its full benefits can be realised.
The approach to the implementation of the performance measures for an IT department falls into three stages:

- Stage 1 – Initial Development Process
- Stage 2 – Continuous Use of the Performance Measures
- Stage 3 – Refreshing and Updating the Performance Measures

Olve *et al* (1999) and Olve and Sjostrand (2002) propose a number of key steps in the implementation process which may be related to the three stages identified above.

6.1 Stage 1 – Initial development process

In this stage it will be necessary to:

- Obtain senior management commitment and support for the project. It is important that the rationale for the development of a performance measurement system is communicated to all stakeholders in order to overcome potential barriers to implementation such as fears about the perceived adverse effects of performance measurement and process improvement;
- Provide information and training in performance measurement and process improvement;
- Form a small project team with appropriate membership in order to take the implementation process forward effectively and obtain ownership from the team and the organisation;
- Decide on a department-wide implementation or pilot the implementation within specific areas of IT. The pilot would have the advantages of allowing the department to become familiar with the performance measurement process, learn from mistakes and gain the confidence and commitment of staff before rolling the process out to the department as a whole;
- Clearly establish links between the development of measures with the corporate and ICT strategies; and
- Integrate reward and recognition schemes with performance improvement as measured by the performance measurement system or BSC.

6.2 Stage 2 – Continuous use of the performance measures

In this stage it will be necessary to:

- Set short-term and long-term goals for each measure which are consistent with the corporate strategy. Equally they must be realistic and attainable, but also challenging;
- Develop flexible and effective systems and procedures to collect information necessary to monitor the performance measurement system. This should draw on readily accessible information and allow the automation of measurement; and
- Focus on a balanced set of measures and explore the causal links between financial and non-financial measures to bring about

process improvements. The emphasis here is on the department or organisation becoming a learning organisation.

6.3 Stage 3 – Refreshing and updating the performance measures

The development of a performance measurement system cannot be regarded as static and must therefore be kept under review as the corporate and ICT strategies evolve and change in response to internal and external drivers.

In summary the implementation process requires initial impetus, commitment, training and co-ordination in order to overcome cultural and organisational barriers; the performance measurement system must then be institutionalised as a process within the organisation so that the organisation's focus is on what it is doing well, what it is not doing so well and what can be improved; and finally the performance measurement system must be kept under review and alive.

7. Monitoring

A number of approaches to monitoring the IT performance measures are considered in this section, but the focus is on putting in place a monitoring process which will lead to improvements in the performance of the IT department.

7.1 Using performance data

It is important to avoid information overload, ensure that information is accurate, up-to-date and credible, use accessible information, present information in an attractive, easy to understand way and to select an appropriate frequency to monitor and act on performance information.

Performance data can be used to:

- Predict trends in order to allow proactive decision-making and actions to be taken. For example, the analysis of the number of enquiries received by the IT Help Desk will in turn act as a predictor of future resources needed to support this function and the most effective deployment of staff resources;
- Correlate results in order to concentrate actions that will give best results. For example, the correlation of information on budgets, timescales or project issues in relation to projects where a formal project management methodology was not used versus projects in which formal project management methodology was used, may give rise to more informed decisions about the appropriate use of project management methodologies; and
- Access to historical performance information will allow year-on-year comparisons and therefore a sounder basis for making informed judgements.

In order to ensure that performance information is actively reviewed and used as a basis for implementing improvements, the use of a more formal monitoring framework could be considered. This would allow the IT Department to ask key questions about its performance:

- What is our current performance?
- How does actual performance compare with the target set?
- What are the main reasons why the target isn't being met?
- What is the plan for corrective action?
- Has the action been taken?
- Does the action have the desired impact on the results of the measures?

Neely *et al* (1996), Neely (1998), Olve *et al* (1999) and Bourne (2000) highlight two frameworks – Ford QOS (Quality Operating system) and Xerox follow-up framework – that could be adopted for use by the IT Department to monitor and continuously improve its performance.

The Ford framework is shown in Figure 2 below and consists of 4 panels:

- Panel 1 – the graph of actual performance against target;
- Panel 2 – a breakdown of that result by the main factors contributing to the result achieving or missing the target;
- Panel 3 – the action planned to improve the performance; and
- Panel 4 – the record of the impact of the action taken.

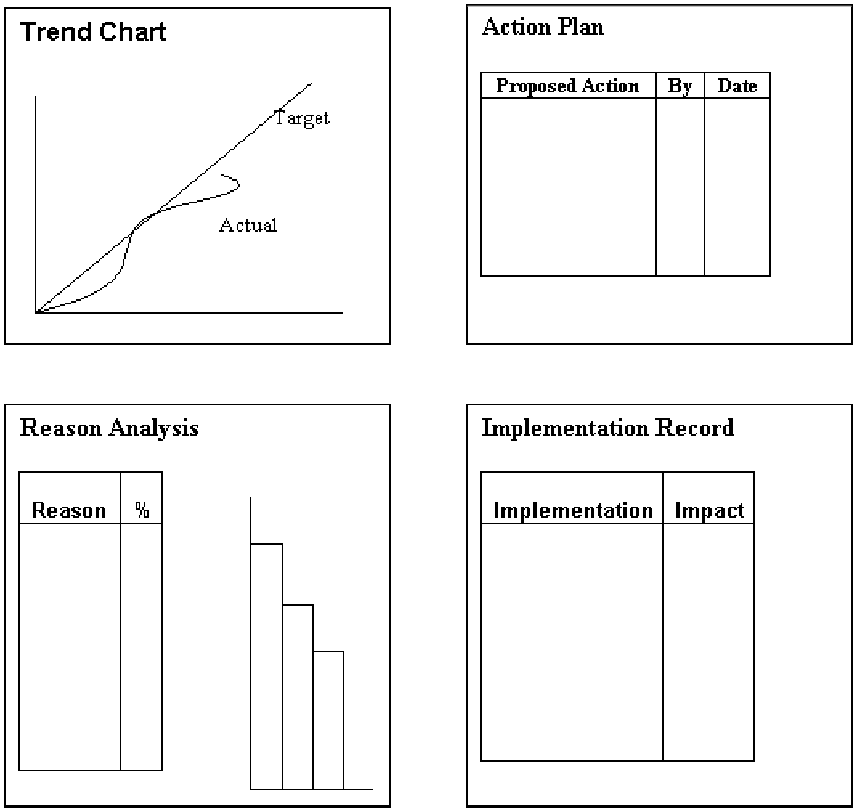

Figure 2: The Ford QOS Measure Visualisation (Adapted from Neely et al, 1996)

The Xerox framework is shown in Figure 3 below and includes:

- An owner, that is an individual responsible for preparing quarterly analysis of the trends, causes, strengths and areas for improvement as well as the action plan;
- A sponsor: in Xerox's case, a main Board Director;
- A desired state, including the results, approach and pervasiveness and a 7 point rating where 7 is "world class performance";
- Performance;
- Causal analysis;
- Strengths;
- Areas for improvement; and
- Detailed action plan.

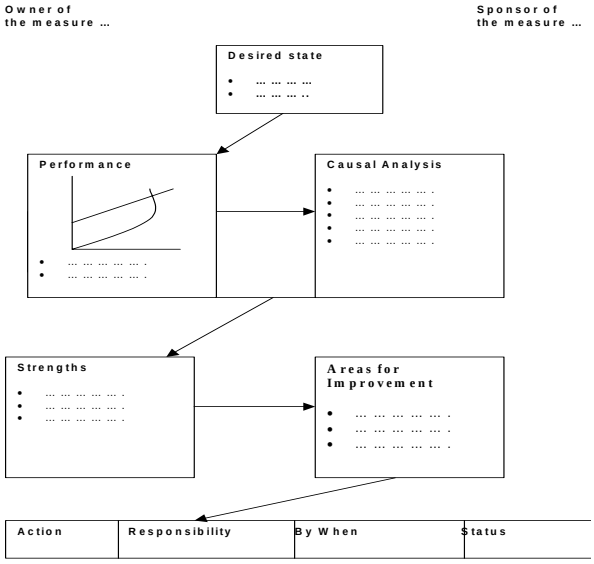

Figure 3: Follow-up of Measures at Xerox (Adapted from Olve *et al*, 1999)

The benefits of these frameworks are that they ensure that action is taken as an outcome of a performance monitoring process and that there is shared ownership of the drive towards improvement.

7.2 Benchmarking performance

Two approaches to benchmarking of the IT performance measures are possible:

- Benchmarking of performance against comparative organisations e.g. using information derived from Gartner (information systems research specialists); and
- Using a "maturity model" which allows an organisation to grade its IT processes in absolute terms from non-existent to optimised (from 0 to 5). This approach is derived from the Maturity Model that the Software Engineering Institute defined for maturity of the software development capability (Paulk *et al*, 1993). Against these levels an organisation can map:
  - The current status of the organisation – where the organisation is today;
  - The current status of the industry (best-in-class) – the comparison;
  - The current status of international standards – additional comparison; and
  - The organisation's strategy for improvement – where the organisation wants to be.

The maturity model is shown in Figure 4 below:

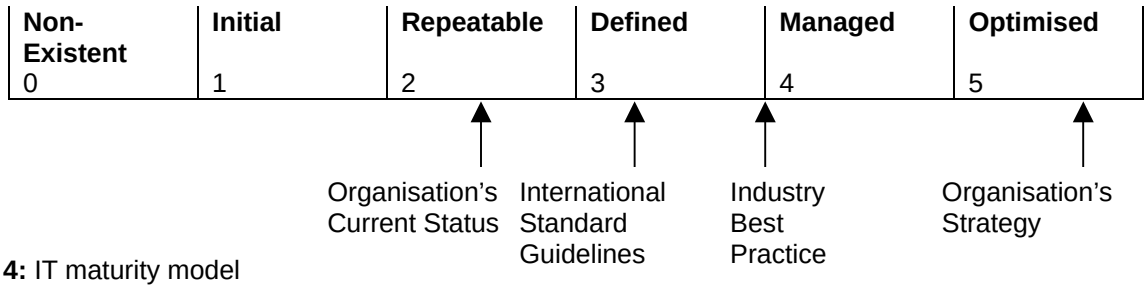

Figure 4: IT maturity model

In this model the scale from "non-existent" to "optimised" can be interpreted as follows:

0	Non-Existent	Management processes are not applied at all
1	Initial	Processes are ad hoc and disorganised
2	Repeatable	Processes follow a regular pattern
3	Defined	Processes are documented and communicated
4	Managed	Processes are monitored and measured
5	Optimised	Best practices are followed and automated

7.3 Communication

In order to ensure that key staff are aware of the monitoring of the IT performance measures and that this is instilled in the culture of the department a number of actions may be taken:

- Display of monitoring information on departmental notice boards and on the intranet;
- Briefings on the results of measures become an integral part of staff meetings and one-to-one reviews of individual objectives. These briefings should clearly demonstrate the impact on the department and on the organisation as a whole of achieving or failing to achieve agreed targets; and
- Briefings to the senior management team on the results of measures.

8. Conclusion

This paper has established a framework and process for the implementation of a performance measurement system in an IT Department within the context of the UK Health Service. Performance measures have been identified, the rationale underpinning the selection of measures has been explained, a Performance Measurement Record Sheet has been used to define how those measures would work in practice, and approaches to implementation, monitoring and reviewing measures have been considered. As an outcome of fully implementing a performance measurement system within an IT department the following benefits may be achieved:

- Through deciding what to measure will encourage the IT team to focus on and clarify what is important for the department within the context of what is important for the organisation as a whole;

- When the measures have been identified and are in place the managers within the IT department will have a means of communicating to the team and senior management a clear framework for working towards the department's goals;
- In turn this will provide a means of influencing behaviour, and ensuring the right things are being done;
- Having established this "route map" the IT department can check on an ongoing basis whether or not objectives are being achieved; and
- Finally, the measurement data can be used to challenge the department's strategy and how well it is integrated with the organisation's mission.

Of paramount importance is to see each measure in terms of what Kaplan and Norton (2000, p69) describe as a "*strategy map*" i.e. each measure as "*embedded … in a chain of cause-and-effect logic that connects the desired outcomes from the strategy with the drivers that will lead to the strategic outcomes".*

References

Bourne, M and Bourne, P *Understanding the Balanced Scorecard in a Week,* Hodder and Stoughton, London (2000).

Department of Health, *Delivering 21st century IT Support for the NHS* (July 2002). Available from: http://www.dh.gov.uk/assetRoot/04/07/16/84/04071684.pdf

Department of Health, *Information for Health: An Information Strategy for the Modern NHS 1998-2005* (September 1998). Available from: www.doh.gov.uk/ipu/strategy/index.htm

Department of Health, *Building the Information Core: Implementing the NHS Plan* (January 2001). Available from: http://www.nhsia.nhs.uk/pdf/info_core.pdf

Department of Health, *The NHS Plan: A Plan for Investment, A Plan for Reform* (July 2000). Available from: http://www.nhs.uk/nationalplan/nhsplan.pdf

Eccles, R G 'The Performance Measurement Manifesto', *Harvard Business Review*, (January-February 1991) pp 131-137.

Kaplan, R and Norton, D P 'The Balanced Scorecard – Measures that Drive Performance', *Harvard Business Review*, (January-February 1992) pp 71-79.

Kaplan, R and Norton, D P 'Using the Balanced Scorecard as a Strategic Management System', *Harvard Business Review*, (January-February 1996a) pp 75-85.

Kaplan, R and Norton, D P *Translating Strategy into Action: The Balanced Scorecard*, Harvard Business School Press, Boston, MA. (1996b).

Kaplan, R and Norton, D P *The Strategy Focused Organisation – How Balanced Scorecard Companies Thrive in the New Business Environment,* Harvard Business School Press, Boston, MA. (2000).

Neely, A *Measuring Business Performance – Why, What, How*, Profile Books Ltd, London (1998).

Neely, A, Richards, H, Mills, J, Platts, K and Bourne, M 'Designing Performance Measures: A Structured Approach', *International Journal of Operations & Production Management,* Vol. 17, No. 11, (1997), pp 1131-1152.

Neely, A D, Mills, J F, Gregory, M J, Richards, A H, Platts, K W, Bourne, M C S *Getting the Measure of Your Business,* Findlay Publications, Holton Kirby (1996).

Olve, N and Sjostrand, A *The Balanced Scorecard,* Capstone Publishing, Oxford, UK (2002).

Olve, N, Roy, J, and Wetter, M *Performance Drivers: A Practical Guide to Using the Balanced Scorecard,* John Wiley & Sons, Chichester, UK (1999).

Paulk, M, Curtis, B, Chrissis, M and Weber, C 'Capability Maturity Model for Software, Version 1.1' *IEEE Software*, Vol. 10, No. 4, (1993), pp 18-27.

Wanless, D *Securing Our Future Health: Taking A Long-Term View*, HM Treasury (2002). Available from: http://www.hmtreasury.gov.uk/Consultations_and_Legislation/wanless/consult_wanless_final.cfm

Table 4: Performance measurement record sheet for IT performance measures

	Project management			IT Help desk		
Measure	*Increased number of projects completed on time and on budget*	*Increased number of projects delivered in accordance with a defined methodology*	*Number of project management training days per project team member*	*Cost per call*	*Increased user satisfaction with the effectiveness and efficiency of the Help Desk*	*Reduced average time to resolve problems*
Purpose	To focus on the control of project timescales and cost	To ensure that significant projects are managed using PRINCE II	To ensure that those involved in managing the project are trained in project management techniques	To minimise the cost per call of the IT Help Desk.	To focus on the efficiency and effectiveness of this process/function.	To focus on the efficiency of this process/function.
Relates to	The need for the IT department to implement projects efficiently	The need for the IT department to implement projects efficiently	The need for the IT department to implement projects efficiently	The need for IT services to be provided cost effectively.	The need for the IT Department to provide effective IT services.	The need for the IT Department to provide effective IT services.
Target	To complete all projects within a 5% tolerance of estimated timescales and cost	100% of significant projects are managed using PRINCE II by (date)	100% of project team members are trained in project techniques by (date)	Cost per call will not exceed £x given agreed service levels.	95% good to high staff satisfaction with IT Help Desk service.	Average time to resolve problem should be no more than 15 minutes.
Formula	Project plan and budget compared with actual timescales and costs	Projects managed using PRINCE II as a % of total number of projects		Overall cost of IT Help Desk function divided by total number of calls.	% breakdown of customer satisfaction survey. Number of complaints received.	Total time per call divided by the total number of calls.
Frequency	Monthly with review at Project Board meetings. Overall annual review of all projects	Quarterly with annual review		Monthly with quarterly review.	Annual review based on survey. Monthly review of complaints.	Monthly with quarterly review.
Who measures?	Project manager	Director of IT		IT Manager.	IT Manager.	IT Manager.
Source of data	Project initiation document, project exception reports and post implementation review	Business Plan and Project initiation documents		Budget Report and IT Help Desk system.	Customer Satisfaction Survey and Complaints Form.	IT Help Desk System
Who acts on the data?	Project manager and Chair of the Project board	Director of IT		IT Manager.	IT Manager.	IT Manager.
What do they do?	Identify causes, manage process and agree corrective action	Establish reason for PRINCE II not being used; agree corrective action with Project manager; organise training		Review volume, type and management of calls with IT Help Desk Officer. Discuss high variances with Director of IT.	Investigate and resolve complaints and causes of below target satisfaction levels. Agree corrective action with Director of IT.	Review data, determine cause of below target performance and agree corrective action with Director of IT.
Notes						

Evaluating the Benefits of Regional Electronic Marketplaces: Assessing the Quality of the REM Success Model

Denise E Gengatharen and Craig Standing
School of Management Information Systems, Edith Cowan University, Australia
d.gengatharen@ecu.edu.au
c.standing@ecu.edu.au

Abstract: A number of regional Internet electronic marketplaces (REMs) have failed or are floundering, partly due to the lack of proper evaluation of their costs and benefits. This paper uses a conceptual REM Success Model to examine the costs and benefits of a REM in Western Australia. The model has been derived from an extension to the Updated DeLone & McClean IS Success Model. The findings from the case study indicate that the REM Success Model, which includes cognisance of SME-profile and motivation of the market maker, allows up-front identification of the costs and benefits to all stakeholders.

Keywords: E-Commerce, Regional Electronic Marketplaces, Small and Medium Enterprises (SMEs), Evaluation of Benefits, REM Success Model

1. Introduction

Internet e-marketplaces continue to receive the attention of the business and academic communities as they strive to understand how to use them to leverage the benefits of the digital economy. Many studies deal with the benefits of e-marketplaces and their critical success factors (Bakos 1998; Braun 2002; Brunn, Jensen, & Skovgaard 2002; Le 2002; Raish 2001). However, there have been few attempts to develop evaluation models with a multi-stakeholder perspective. This is because e-marketplace models are evolving and while there are common benefits, some can be differentiated on the basis of e-marketplace structure and ownership models. One factor that predicates the need for an e-marketplace evaluation-framework is the number of government-sponsored regional e-marketplaces (REMs) that are being considered and established (for example the London Marketplace, the Roses Marketplace, Essex marketplace and the Slough IDeA:marketplace in the United Kingdom) despite past failures (Tasmania Business Online e-marketplace (Hayes 2004), Food Connect Australia (Wilkins, Swatman & Castleman 2003) and Bizewest (Western Melbourne Regional Economic Development Organization (WREDO) 2003) in Australia).

The growth of REMs is a result of direct intervention by local governments who view them as a means to promote regional growth and encourage uptake of e-commerce by SMEs. This is one response to research showing that only a small proportion of SMEs were realising substantial benefits from the Internet because of lack of technological expertise and uncertainty about the benefits offered by e-commerce (Daniel, Wilson, & Myers, 2002; Poon & Swatman, 1997; Walczuch, Van Braven, & Lundgren, 2000). However, governments need to balance the desire for regional economic growth via REMs with a proper evaluation of the costs and benefits associated with developing and managing them. Failure to do so could result in wasted public funds and losses to the SMEs involved. Tonkin (2003) likens the use of government procurement REMs to achieve a broad range of policy objectives in the absence of adequate evaluations, as little more than an act of blind faith.

While existing IS success models like the Updated DeLone & McLean IS Success Model (DeLone & McLean, 2003) (hereafter referred to as the Updated D&M IS Success Model) can be used to measure the success of e-commerce information systems, specific benefit-evaluation frameworks for REMs can provide existing and potential market makers with a clear idea of the costs and benefits to be considered. This would be particularly useful given the complexities involved in establishing and maintaining REMs and as research indicates that generally the investments needed to create e-marketplaces are very high (Brunn et al. 2002).

This paper uses the REM Success Model (Gengatharen & Standing, 2003a) to evaluate a REM in Western Australia. The strengths and limitations of the model are discussed in relation to how it can be used to help market makers and participants recognise the costs and benefits associated with REMs.

2. REM success model

Gengatharen & Standing (2003a) propose a conceptual model to evaluate government-sponsored REMs (figure 1). The model, an extension of the Updated D&M IS Success Model, takes a longitudinal approach and considers the context of the evaluation. This includes analysing the motives for the development of the REM using a stakeholder perspective and assessing the profile of the regional SME sector.

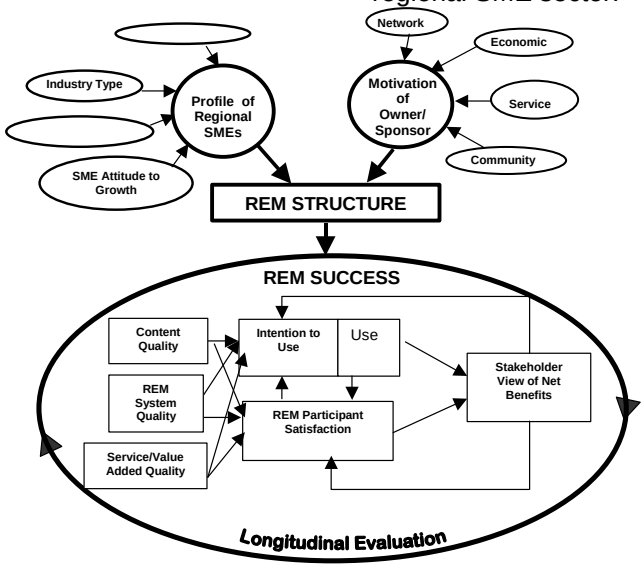

Figure 1: REM Success Model (Gengatharen & Standing 2003a)

In figure 1, the context of the evaluation is illustrated in two areas. Firstly, the benefits accruing from the REM depend on the structure of the marketplace and its ownership model, which in turn are dependent upon the market maker's motivation for REM development and regional SME profile (see Gengatharen & Standing (2003b) for a discussion on market-makers' motives). Secondly, evaluation of benefits depends on which stakeholder's view is being considered i.e. buyers, sellers, owners, intermediaries, other stakeholders or the region. Quality of the content, system and value-added services of the REM determine the participants' intention to use, their actual use and satisfaction with the REM. The more satisfied they are with the REM, the more participants will use it and this determines the benefits that they obtain from using it. The benefits then reinforce the participants' intention to use, their actual use and satisfaction with the REM. The longitudinal aspect of the REM Success Model considers the evolving nature of REMs in order to identify the benefits that may be experienced further along the REM maturity curve.

2.1 Costs

By including market maker's motivation and SME profile within the evaluation model, the market maker is provided with an idea of the costs associated with addressing these issues in REM development. For example, the network motive will require the market maker to consider the cost of creating liquidity on the REM. This could involve low participation fees and low cost training or advice to SMEs on e-commerce and REM participation. Grewal, Comer & Mehta (2001) hypothesised that organizational motivation and ability are important determinants of e-marketplace participation. There is a need to identify key players, determine REM features that would motivate them to join and help them to migrate their transactions to the e-marketplace to create early liquidity (Grewal et al., 2001). This has implications for the level of funding needed to develop and maintain the REM. The economic motive would involve reducing costs for participants and improving efficiency possibly through integration of the REM system with that of participants. Table 1 summarises the costs that could be associated with a REM. It has been derived from the following sources: the extant literature on e-marketplaces, IT/e-Commerce evaluation and SME uptake of e-Commerce and our on-going involvement with a REM in Western Australia.

Table 1: Summarised Costs of REMs

	Owners	**Sellers**	**Buyers**	**Intermediaries**
Direct Costs	Hardware/Software/ Network set-up & maintenance costs User Training & advice costs Marketing & Administration Overheads Provision of value added/ trust/ security services Maintaining parallel systems/ Integration costs	Participation fee Transaction fee Flyer fee Hardware/ software/ network & maintenance costs Administration costs Integration costs (business processes) Training costs	Hardware/ software/ network & maintenance costs Administration costs Integration costs (business processes) Training costs	Advertising Hardware/ software/ network costs Training costs
Indirect Costs	Opportunity costs & Time			

2.2 Benefits

To determine the benefits of a government-sponsored REM, Gengatharen & Standing expanded the success metrics from the Updated D&M IS Success model to include benefits identified in the literature on e-commerce (Molla & Licker, 2001; Zhuang & Lederer, 2003, p. 71) and e-marketplaces (Bakos, 1991; Benjamin & Wigand, 1995; Kaplan & Sawhney, 2000; Standing & Stockdale, 2001). The context of the evaluation was also considered by expanding the net benefits category according to individual/organizational/ industry benefits and regional benefits. Table 2 summarises the REM success metrics.

Table 2: REM success metrics

Content Quality	**REM System Quality**	**Service/Value Added Quality**
Quality of content Accuracy & currency of content Security, Privacy, Authenticity Comprehensiveness Timeliness Relevance Completeness *Quality of content presentation* Effective/meaningful organization of content Navigation techniques Logical structure of content Personalization	Reliability of software/network Accuracy of system Flexibility/Adaptability Ease of Use Online response time & page loading speed System architecture Visual Appearance Convenience of accessibility Market reach Integration with participants' systems	Trust Neutrality of market-maker Site intelligence (CRM) Feedback mechanisms Relevant search facilities Calculators Tracking capabilities Helpdesk/Set-up help/Advice Account maintenance Training (in conversion to e-business & system use) FAQ's
Use	**REM Participant Satisfaction**	
Informational Number of Hits/Visits Transactional Number of Transactions *Community* Presence of & participation in networks	Overall satisfaction Satisfaction with REM Offerings (Repeat visits, repeat transactions, use of networks and use of feedback mechanisms) Satisfaction with governance structure	
Stakeholder Net Benefits		
Industry/Organizational/ Individual Impact	*Transactional Benefits* Lower transaction, staff, communication, search, marketing/advertising, inventory-holding costs Cheaper prices Lower inventory holdings *Productivity gains* Time savings Process efficiencies Wider market reach	Strategic Benefits Gains from Network Externalities & Collaboration Image/Legitimacy Improved market share Improved communications with customers (CRM), suppliers & employees Improved decision making process
Regional Impact	Economic Benefits Attractive location for business Attractive to skilled labour Efficient show-casing of regional offerings Reduced communication costs Increased productivity	Community/Strategic Benefits Collaboration/Partnerships Cooperation Increased level of on-line participation Become a knowledge region Narrow/Close digital divide Relationships with other markets

3. Research design

The previous sections of the paper have highlighted the complexity of evaluating the success of REMs. We want to determine if the constructs identified in the REM Success Model (figure 1) provide a comprehensive framework for identifying factors related to: The motivation of the owner/sponsor, the SME profile, and REM measures of success (including costs and benefits). Therefore, our major research question is:

> *"Is the* REM Success Model *a comprehensive framework for evaluating the success of a REM?"*

To answer this question and evaluate the REM Success Model we have used a case study approach. Case studies can be used to evaluate theory (Yin 1994, p.38) to determine if the constructs are valid. They are useful when a phenomenon is broad and complex and when an holistic in-depth investigation is needed (Dube & Pare 2003).

For the case study, background data on the REM was collected from official documents, through face-to-face discussions, e-mail correspondence and attendance at meetings with the REM owners. To date, in-depth semi-structured interviews of between 1 to 1½ hours duration have been conducted with nine owner representatives and seven SME participants as the REM has only been actively promoted since the beginning of 2004. The owner representatives were the REM project manager, the economic development managers and heads of purchasing of the towns and REM board members representing the towns and business associations. The interview schedules were designed around the success metrics in the REM Success Model. For the SMEs, a preliminary questionnaire relating to company demographics (e.g. nature of business, size of company, date of incorporation, number of employees, annual turnover, etc.) and REM usage (e.g. date of joining, costs of participation, levels of IT competence, REM usage, monthly business volume via the REM etc.) was distributed and collected back before the interviews. Responses were discussed with participants during the interview. The interview also consisted of questions relating to the participants' perception of the REM content, system and service quality, as well as actual or perceived benefits of their participation. For owner representative interviews, statistics on REM usage, set-up and ongoing costs, REM income according to type etc. were collected prior to the interviews. The data was discussed during the interviews together with questions about perceptions of the REM and its offerings and actual and perceptual benefits to themselves, SMEs and the region.

The data collected during the study was transcribed and stored in a qualitative research software package. The constructs from the REM Success Model were used as tags and a structural framework developed. Findings indicate that there appears to have been a lack of understanding of the actual costs and benefits of the REM by owners and participants. The use of the REM Success Model and its success metrics has helped owners and participants obtain a clearer idea of these costs and benefits. Data gathering from interviews will continue and a second round of data collection is planned in six months to a year to cater for the longitudinal aspect of the model. The REM Success Model will also be used in cross case analysis.

4. Case study

TwinTowns (a pseudonym) is a web portal incorporating community content, a business directory and a REM set up by two neighbouring towns in Western Australia. For the purposes of funding and bargaining power, the owner/sponsor group of the portal was widened to include the local councils and business associations of the two towns, and two local higher education institutes (HEIs). The REM is intended to be an electronic gateway to access and interact with local players in the area. Local registered suppliers pay an annual fee of A$ 199 and an additional fee of A$ 99 for a business flyer page. Community groups can have content hosted free of charge on the portal.

The idea of the portal was first conceived in 1999 and with a small seed-funding grant from the state government, a demonstration site was created and used to promote the project and obtain further funding. This was obtained from federal and state government sources and the two towns, their business associations and one of the HEIs. It was intended that after the initial funding was exhausted, TwinTowns would be self-sustainable using income from participation fees, advertising fees and sponsorship. Development of the portal was outsourced in early 2002 and in December 2002, the portal went live with a 'soft-launch'.

The presence of the REM within the portal is intended to provide SMEs in the region with a low cost, low technology-compliance introduction to e-marketplace trading. The business directory in the REM will allow consumers to source products from local suppliers. The REM also provides an e-procurement mechanism for local and other buyers to request for and receive quotations from registered local SME suppliers and to place orders with them. The key motivations for developing the REM are increasing e-commerce adoption by SMEs, improving business efficiency and increasing trade within the locality, expansion into new markets and development of the region generally.

To date, TwinTowns has a total of 157registered REM participants and 37 community groups while close to A$ 400,000 has been expended on the project. A shortfall in expected funding from government sources, a huge underestimation of the resources required and problems with the software vendor have seen the motive of TwinTowns change to pure survival. Only since the beginning of 2004, with sponsorship funds from a real estate developer, a sales and marketing plan has been put in place. This has seen the number of participants increase by 50 percent. The REM is now being sold on the basis that only *discretionary purchasing* of the major regional buyers will be channelled via the REM. This is because the e-procurement systems of the major regional buyers cannot be integrated with the REM system.

5. Applying the REM success model

The following is a discussion of the application of the REM Success Model to TwinTowns. Although only 7 registered SMEs have been interviewed so far, it is nevertheless a good representation as requests for quotes (RFQs) have only been issued since early 2004.

5.1 Market- Maker's motivation

In TwinTowns, the motivation to stimulate regional development by encouraging SMEs to participate in the REM via a 'buy local' policy resulted in the horizontal nature of the market and the RFQ/Ordering mechanism. Entry and technology-compliance costs for SMEs were kept low but lack of resources meant that the portal is currently trying to survive by promoting the REM, while the community side of the portal lies dormant. At this stage, only the discretionary purchasing of the two towns is beginning to be routed through the REM. It is interesting to note that while the majority of the owner representatives interviewed stated that the original motive of the REM was to "promote regional economic growth" by using the REM to stimulate a "buy local policy" and create "local opportunities", no benchmark figures were available to determine how much spending by the two towns was being channelled into the local market in the first place. Those figures are now being compiled. Another interesting point is that although economic development was claimed as the motive, the original business plan was based on the REM being self-sustainable within two years of operation. This was attributed to the "hubris surrounding the dot.com hype" and "an incrementalist fund-sourcing strategy" where additional funding will be sought after results can be demonstrated.

5.2 Regional SME-profile

SME-profile will influence the structure of the REM and inter-alia the type, level and timing of benefits that can be delivered. While TwinTowns was envisaged initially as a B2B, B2C and B2G REM, local suppliers comprised mainly micro businesses with fewer that 10 employees. All seven SMEs interviewed were in this category and most did not consider using the REM for purchases. One saw it as an avenue for B2C commerce given that the community part of the portal was in the initial plan. Of the remaining 6 SMEs, 5 saw the REM as an opportunity to get a slice of local government business and "hopefully some from the other major local buyers". However, there was no attempt by the proponents of the REM to determine how the REM system could be integrated with the e-procurement systems of the local major buyers, although some of the owner representatives indicated that the IT department of one of the towns did try to voice their concerns.

A region having strong offline SME networks or alliances may provide impetus for early on-line collaboration in REMs. In TwinTowns, the local business associations represent less than twenty percent of the businesses in the region and collaborative-commerce is not something the SMEs are familiar with. Although complaining about the local governments' bulk discount purchases from large non-local companies, when the question of collaboration and aggregation of supplies by small local suppliers was raised, one SME replied that it was "pie in the sky stuff" and "if it was me & I had to deal with 5 or 6 other people to get a contract, the hassle that goes with it will be too great".

Regional SME attitude towards growth will also have a bearing on the structure of the market, its features offered and the benefits to be gained. If the majority of SMEs are lifestyle SMEs (Jeffcoate, Chappell, & Feindt 2000), the owners of the REM may choose to consider an e-marketplace model that reflects the SME owners' strategy for business growth (Levy & Powell 2003) and market the REM as such.

5.3 Content quality

A number of studies indicate that content quality can affect the satisfaction of web site users (Molla & Licker 2001, p. 138). In the case of TwinTowns, there has been little maintenance of the database of registered suppliers. For example, when the authors tried to send an email to a REM participant, the email was no longer valid. Of the seven SMEs interviewed, four indicated they would not use the REM to look for accurate details of local suppliers, while the others could not comment as they only went into the REM to look at their own listings "some time ago". On the question of relevance of the portal content, 3 SMEs said they had "no idea" but one qualified the statement by saying that "at that time (more than a year ago) it seemed OK". Of the remaining 4 SMEs interviewed, one thought the information was relevant (this SME has obtained an RFQ via the REM and has become a supplier to one of the major local buyers), two felt that there was insufficient content and the fourth, a newly incorporated IT services business, felt that the content was "not useful" and appeared "unprofessional". Although all the SMEs felt that privacy, security and authenticity of the content were important none felt that these aspects had been tested given the very low level of activity on the REM.

5.4 REM system quality

For many of the SMEs registered with TwinTowns this is their first encounter with e-marketplace participation and system performance is critical to their satisfaction and continued use. The majority of the SMEs interviewed felt that the REM system was either unreliable or they were unable to comment as they had not used it for some time. However they felt that reliability of the REM system would be vital if they were actively transacting through it.

The extent to which the REM system can be integrated with participants' systems is also a measure of the system's quality. This is where the longitudinal approach to benefits evaluation comes into play. In the short term the need for integration with the SMEs' systems may not be critical; it could assume a bigger role as SMEs move further along the e-commerce adoption ladder towards the theoretical end-point of becoming e-businesses (Commission of the European Communities, 2003; DTI, 2001). The majority of SMEs interviewed have not given much thought to integration of the REM system with theirs, but agreed that it could be useful. The longitudinal evaluation is also important as the REM may evolve and its motives could change. Although the success of TwinTowns now appears to hinge on B2G commerce, there appears to be no consideration by the towns of their costs and benefits of procuring through the REM. For example, the REM system is not integrated with the e-procurement systems of the towns and they are running two separate systems for their procurement needs, which could impact on efficiency. Although the REM will be used for the discretionary purchasing of the towns (currently each prospective supplier is manually contacted for quotes), any efficiency savings to them have not been factored into the return on investment of the REM, despite claiming that one of the benefits was the "time and efficiency savings in the process of discretionary purchasing". In terms of ease of use, all SMEs interviewed rated the REM positively and felt it would be a convenient tool, if there was critical mass.

5.5 Service/value added quality

The service that the REM performs in providing SMEs with e-business training and REM usage, in connecting them with trusted providers of intermediary services (Lenz, Zimmerman, & Heitman 2002) and in offering adequate technical help will have a bearing on the REM participants' satisfaction and use of the REM. With the TwinTowns REM, although there were plans to provide bundled services to participants by tying up with service providers in the region, these have now been abandoned as the REM struggles to survive. The issue of trust in a government-sponsored REM is one worth mentioning. In TwinTowns, the initial move to register buyers on the REM provided some mechanism of authenticity or trust. However this has now been discontinued, as it is too costly. Feedback can also affect the success of a REM. In TwinTowns, there has been no feedback from the system to participants who have not been sent RFQs, to the extent that the majority of them doubted that the system was working. There are plans now to send out fortnightly or

monthly listings of RFQs issued via the REM and reports on REM success stories. Some of the SMEs indicated that knowing who the tender was awarded to could also allow them the opportunity of approaching the awardee for sub-contract work, while a number of the SMEs felt that a feedback system would provide some transparency of the tendering process.

5.6 Use

The use of the REM covers activities like obtaining information (measurable by the number of hits), performing transactions (measurable by number of transactions) and participating in networks (measurable by the presence of and participation in clusters and forums). While there has been some argument over 'use' as a measurement of success in IS evaluation, use of a government-sponsored REM is at the discretion of the participant and is therefore a good indicator of REM success. Only 3 of the seven SMEs interviewed have continued to use the REM as they have received RFQs via the system. However, they only use it for responding to RFQs and not for other purposes.

5.7 REM participant satisfaction

This is a measure of how participants feel about all aspects of the REM. Feedback mechanisms can be used to measure satisfaction, as can indicators like repeat visits and transactions and discussion forums. While all the SMEs felt that the concept was good, they were not satisfied due to the low volume of business, the lack of promotion of the REM to the wider public and the lack of feedback.

5.8 Net benefits

This is the ultimate measure of REM success, as it will determine the benefits and negative effects that the REM will have for all stakeholders. The longitudinal approach to evaluation is important as some of the benefits may only be realised in the longer term. Some of the REM net benefits measurements would still be the ones "developed and tested for IS investments in general" (DeLone & McLean 2003, p. 25). These would measure the benefits that fall under the individual and organizational net transactional benefits categories. While the majority of SMEs have not seen any benefits from their participation in TwinTowns yet, they expect that at the very least, if the REM worked as envisaged, they would save on advertising and communication costs. In terms of net strategic benefits, SMEs felt that the REM offered networking opportunities and entry into the "loop of the local major players". Many of the economic and strategic benefits of the REM for the region will only be realised in the longer term. However the REM Success Model is still useful as it provides an idea of the data that needs to be collected now as benchmarks for future measurement and evaluation. In terms of collaboration and partnerships, for a start, some of the SMEs and owner representatives felt that the project has brought the major stakeholders in the region together in a collaborative effort to improve e-commerce uptake in the region. Not surprisingly all sixteen interviewees felt that the REM was not successful because of low usage resulting from lack of promotion, funding shortfalls and technical problems.

6. Conclusion

The number of REMs being developed for SMEs, often where the market makers and/or participants do not have a full understanding of the costs and benefits associated with them, predicates the need for an evaluation framework that can encompass a more holistic approach to e-marketplace evaluation. The REM Success Model is a useful evaluation framework that can be used to design instruments to measure the costs and benefits associated with REMs. By using the extant literature on e-marketplaces, IT/e-Commerce evaluation and data from an actual REM, a list of costs was determined according to each type of stakeholder in the REM. Given the wider socio-economic objectives of REMs, the success metrics identified in the framework were not only direct benefits, but indirect ones as well. According to Bakos (1991), e-marketplaces are socially desirable when net welfare gains are greater than development and operating costs. By identifying the costs and benefits to all stakeholders, the REM Success Model helps determine the baseline data that needs to be collected against which to measure success or lack thereof. The REM Success Model is both descriptive and prescriptive as it illustrates the success factors of the REM while also indicating how they can be measured. Future research will include testing the model on other REMs.

References

Bakos, Y 'A Strategic Analysis of Electronic Marketplaces' *MIS Quarterly*, Vol 15 No 3 (1991) pp 295-310.

Bakos, Y 'The emerging role of electronic marketplaces on the Internet'

Communications of the ACM, Vol 41 No 8 (1998) pp 49-55.

Benjamin, R and Wigand, R 'Electronic markets and virtual value chains on the information superhighway' *Sloan Management Review*, Vol 36 No 2 (1995) pp 62-72.

Braun, P 'Digital Knowledge Networks: Linking Communities of Practice with Innovation' *Journal of Business Strategies*, Vol 19 No 1 (2002) pp 43-54.

Brunn, P, Jensen, M and Skovgaard, J 'eMarketplaces: Crafting A Winning Strategy' *European Management Journal*, Vol 20 No 3 (2002) pp 286-298.

Commission of the European Communities, 2003. Adapting e-business policies in a changing environment: The lessons of the Go Digital initiative and the challenges ahead [Online]. Available from: http://www.eu.nl/comm/enterprise/ict/policy/doc/com_2003_148_en.pdf [Accessed: 21May 2003].

Daniel, E, Wilson, H and Myers, A 'Adoption of e-commerce by SME's in the UK: towards a stage model' *International Small Business Journal*, Vol 20 No 3 (2002) pp 253-268.

DeLone, W H and McLean E R. 'The DeLone and McLean Model of Information Systems Success: A Ten-Year Update' *Journal of Management Information Systems*, Vol 19 No 4 (2003) pp 9-30.

DTI. (2001). Business in the Information Age. International Benchmarking Report [Online]. Available from: http://www.ukonlineforbusiness.gov.uk/main/resources/publication-htm/bench2001.htm [Accessed: 7 May 2003].

Dube, L. & Pare, G. (2003). 'Rigor in Information Systems Positivist case Research: Current Practices, Trends, and Recommendations' *MISQ*, Vol. 27 No 4 (2003) pp. 597-635.

Gengatharen, D. E., & Standing, C. (2003a). 'A Proposed Model to Evaluate the Benefits of Government-Sponsored Regional Electronic Marketplaces for SMEs: Extending the Updated DeLone & McLean IS Success Model'. In Proceedings of the 4th International We-B Conference (Croft, L., Lachowicz, E. & Knight, S. Eds.) We-B Centre, Edith Cowan University, Perth, Australia.

Gengatharen, D E and Standing, C (2003b). 'A Conceptual Framework to Support the Development of Government-Sponsored Community-Portal Regional Electronic Marketplaces for SMEs: A Stage Approach' In Proceedings of the 14th Australasian Conference on Information Systems (ACIS) (Lethbridge, N. Ed.), We-B Centre, Edith Cowan University, Perth, Australia.

Grewal, R, Comer, J M and Mehta, R 'An investigation into the antecedents of organizational participation in business-to-business electronic markets' *Journal of Marketing*, Vol 65 No 3 (2001) pp 17-33.

Hayes, S. (2004, June 1). Island's $1.8 e-mart crashes. australian IT.com au. [Online]. Available from http://australianit.news.com.au/com…/0,7208,9706933%5E16123%5E%5Enbv%5E,00.htm [Accessed: 2 June 2004].

Jeffcoate, J, Chappell, C and Feindt, S 'Attitudes towards Process Improvement among SMEs Involved in E-commerce' *Knowledge and Process Management*, Vol 7 No 3 (2000) pp 187-195.

Kaplan, S and Sawhney M 'E-Hubs: The new B2B Marketplaces' *Harvard Business Review*, Vol 78 No 3 (2000) pp 97-103.

Le, T T 'Pathways to Leadership for Business-to-Business Electronic Marketplaces' *Electronic Markets*, Vol 12 No 2 (2002) pp 112-119.

Lenz, M, Zimmerman, H-D and Heitman, M. 'Strategic Partnerships and Competitiveness of Business-to-Business E-Marketplaces: Preliminary Evidence from Europe' *Electronic Markets*, Vol 12 No 2 (2002) pp 100-111.

Levy, M and Powell, P 'Exploring SME Internet Adoption: Towards a Contingent Model' *Electronic Markets*, Vol 13 No 2 (2003) pp 173-181.

Molla, A and Licker, P S 'E-Commerce Systems Success: An Attempt to Extend and Respecify the Delone and Maclean Model of IS Success' *Journal of Electronic Commerce Research*, Vol 2 No 4 (2001) [Online]. Available from: http://www.csulb.edu/web/journals/jecr/ [Accessed: 21 March 2003].

Poon, S and Swatman, P M C 'Internet-Based Small Business Communication: Seven Australian Cases' *Electronic Markets*, Vol 7No2 (1997) pp 15-21.

Raish, W D *The eMarketplace. Strategies for Success in B2B Ecommerce*, McGraw-Hill, New York (2001).

Standing, C and Stockdale R (2001). 'Evaluating the benefits of Electronic Marketplaces' In Proceedings of the 2nd International We-B Conference (pp. 118-127). We-B Centre, School of MIS, Edith Cowan University Perth, Australia.

Tonkin, C., 2003. eProcurement in the Public Sector: Story, Myth and Legend. Working Paper. In: The Policy Institute, Trinity College, Dublin. [Online]. Available from: http://www.policyinstitute.tcd.ie/working_papers/PIWPO7%20-%20Tonkin.pdf [Accessed: 27 January 2004].

Western Melbourne Regional Economic Development Organisation (WREDO)., 2003. bizEwest B2B Portal. [Online] Available from: http://www_melbwest.com.au/wredo/projects/bizewest.html [Accessed: 19 January 2004].

Wilkins, L, Swatman, P M C and Castleman, T, 2003. Electronic Markets and Service Delivery: Governance and Related Competencies in Virtual Environments. In: School Working Papers - Series 2003, SWP 2003/02. School of Information Systems, Deakin University, Melbourne, Australia. [Online]. Available from: http://www.deakin.edu.edu.au/infosys/research/working_paper.htm [Accessed: 13 August, 2003]

Walczuch, R, Van Braven, G and Lundgren, H. 'Internet Adoption Barriers for Small Firms in The Netherlands' *European Management Journal*, Vol 18 No 5 (2000) pp 561-572.

Yin, R. K.). *Case Study Research Design and Methods* (2nd Ed). Sage Publications, Thousand Oaks (1994).

Zhuang, Y and Lederer, A L 'An Instrument for Measuring the Business Benefits of E-Commerce Retailing' International *Journal of Electronic Commerce*, Vol 7 No 3 (2003) pp 65-99

.

Questionnaire Based Usability Evaluation of Hospital Information Systems

Kai-Christoph Hamborg[1], Brigitte Vehse[1], Hans-Bernd Bludau[2]
[1]University of Osnabrück, Germany
[2]University of Heidelberg, Germany
khamborg@uni-osnabrueck.de
bvehse@yahoo.de
hans-bernd.bludau@med.uni-heidelberg.de

Abstract: The widespread distribution of HIS requires professional evaluation techniques. In this study we present a usability questionnaire called IsoMetrics which is based on the international standard ISO 9241 Part 10. The questionnaire was applied to assess the usability of a Hospital Information System. The equivalence of the online and a paper-and-pencil format of the questionnaire were investigated. The results show that the different formats do not affect the subject's ratings. IsoMetrics was proven to be a reliable technique for software evaluation in the field of hospital information systems supporting usability screenings in large organisations.

Keywords: Evaluation, usability, ISO 9241 Part 10, Hospital Information Systems (HIS), online questionnaire

1. Introduction

The widespread distribution of Hospital Information Systems (HIS) in healthcare institutions requires professional evaluation to assess the practical usefulness of these applications. So far, evaluations of HIS have been undertaken focussing mainly on financial aspects or considering the patients interests. A major aspect has been neglected: The user! Nurses, physicians and other healthcare employees, working with the software, spend a lot of time each day by filling in forms, reviewing medical inspection results and handling an amount of information for administration needs.

The usability of a product is considered as a precondition of the usefulness of an application (Nielsen, 1993). It is defined with respect to " the extent to which the product can be used by specific users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specific context of use." (ISO 9241 Part 11, 1998). Unfortunately today not many applications fulfill this demand, and thus cause errors, trouble and stress as well as high costs on the part of the users and the organisation (Landauer, 1995).

Usability evaluation aims at identifying strengths and weaknesses of an application and gives hints for improving its usability. There is a multitude of methods for the purpose of software evaluation (Gediga, Hamborg & Düntsch, 2002). Questionnaires are well suited for the summative evaluation of software applications, especially in larger organisations like hospitals, public administrations etc. They are economic evaluation techniques which can be applied to a larger number of users at the same time with comparatively small financial effort.
In this paper the IsoMetrics Inventory (Gediga, Hamborg & Düntsch, 1999) for summative and formative evaluation of software usability will be presented. Its application in an evaluation study concerned with the usability of a HIS is demonstrated. In this study, we established an online version of the questionnaire, aiming at reducing efforts and at speeding up recurrent surveys and consecutive data evaluation. The equivalence of the paper-and-pencil and the online format is examined as well as the reliability of the questionnaire in the application area HIS.

2. Research methodology

The IsoMetrics questionnaire will be presented in the context of an evaluation study which was conducted at the University Hospital of Heidelberg, Department of Internal Medicine.

2.1 Material and methods

2.1.1 The IsoMetrics questionnaire

The IsoMetrics usability inventory (Gediga, Hamborg & Düntsch, 1999) provides a user-oriented, summative as well as formative approach to software evaluation on the basis of ISO 9241 Part 10. While summative evaluation is typically quantitative and located at the end of a development process, using numeric scores to assess the usability of an application, formative evaluation provides (often qualitative) information about weaknesses useful in improving the usability of a software system during the engineering life cycle or before further development. Accordingly there are two versions of

IsoMetrics, both based on the same items: IsoMetrics^S (short) supports summative evaluation, whereas IsoMetrics^L (long) is best suited for formative evaluation purposes. The inventory is available as English and German language version and can be administered by either paper and pencil or an online (inter-/intranet) version. The current version of IsoMetrics (2.04 german/2.01 english) comprises 75 items operationalising the seven design principles of the international standard ISO 9241 Part 10. ISO 9241 formulates „Ergonomic requirements for office work with visual display terminals (VDTs)" and provides guidance for the ergonomic design of interactive software. It comprises 17 different parts, whereas Part 10 covers seven principles for dialog design (s. Table 1).

Table 1: Dialogue Principles according to ISO 9241 Part 10 (translated from the german version by the authors).

Suitability for the task	A dialogue is suitable, if it supports the user to realise his tasks effectively and efficiently. Only those parts of the software are presented, which are necessary to fulfil the task.
Self-descriptiveness	A dialogue is self-descriptive, if every step is understandable in an intuitive way, or, in case of mistakes supported by immediate feedback. Further, an adequate support should be offered on demand.
Controllability	A dialogue is controllable, if the user is able to start the sequence and influence its direction as well as speed till he reached his aim.
Conformity with user expectations	A dialogue is conform with the user expectations, if it is consistent, complying with the characteristics of the user, e.g. taking into account the knowledge of the user in that special working area, accounting education and experience as well as general acknowledged conventions.
Error tolerance	A dialogue is error tolerant, if the intended deliverable is reached with no or just minimal additional effort despite of obvious faulty steering or wrong input.
Suitability for individualisation	A dialogue is suitable for individualisation, if the system allows customising according to the task as well as regarding the individual capabilities and preferences of a user.
Suitability for learning	A dialogue supports the suitability of learning, if the user is accompanied through different states of his learning process and the effort for learning is as low as possible.

The statement of each item of the IsoMetrics^S Questionnaire is assessed on a five point rating scale starting from 1 ("predominantly disagree") to 5 ("predominantly agree"). A further category ("no opinion") is offered to reduce arbitrary answers.

IsoMetrics consists of the same items as IsoMetrics^S and uses the same rating procedure. Additionally, each user is asked to give a second rating, based upon the request "Please rate the importance of the above item in terms of supporting your general impression of the software." This rating ranges from 1 ("unimportant") to 5 ("important"). A further "no opinion" category may also be selected. In this way, each item is supplied with a weighting index. To evoke information about malfunctions and weak points of the system under study, the question "Can you give a concrete example where you can (not) agree with the above statement?" is posed. This gives users the opportunity to report problems with the software, which they attribute to the actual usability item.

IsoMetrics has proved its practicability in software development projects and field studies. Given ten evaluating users, IsoMetrics^L evokes approximately one hundred remarks addressing weak-points of a given software. Its reliability was examined and confirmed for each of the seven design principles (Gediga, Hamborg & Düntsch, 1999, Gruber, 2000). In order to validate the IsoMetrics inventory, the scale means of five different software systems were analysed and compared. It could be shown that programs with different ergonomic qualities were discriminated by the corresponding scales (Gediga, Hamborg & Düntsch, 1999).

2.1.2 Software

The software examined, "IS-H*MED" release 4.63B by T-Systems, Austria is based on the IS-H solution by SAP, Germany. It is mainly table-oriented software with a broad range of functions:

- *Creation of discharge letter*: A discharge letter is most often dictated on tape by a physician and afterwards typed by a secretary. Proof-reading and corrections are realised online, using a MS Word plug-in.
- *View of laboratory and diagnostic findings*: For each patient, an overview of existing laboratory and diagnostic findings is available. A list of the findings permits the physician a detailed look.
- *Documentation of diagnostic finding*: In-patients can be selected by a physician from a listing of the patients to feed in diagnostic findings. The ICD10-Code of the diagnoses might be entered directly – or with the help of a plug-in called KODIP. This plug-in covers the complete ICD-10 via a thesaurus and offers additional information about the grouping accounting rules etc.
- *Diagnose related grouping*: After the individual diagnostic findings and resulting

medical procedures (e.g. a heart catheter) are entered into the computer, a calculation of the Diagnose Related Group's (DRG) may be accomplished.

- *Order of medical examinations* supports the electronical ordering of medical examinations for a patient.
- *Documentation of physical examinations*: This function allows to document the results of an inspection, e.g. an ultrasound sonic examination, or a radiology report. The reports are mainly written with help of a MS Word plug-in.
- *Nursing category*: A staffing calculation methodology derived from the traditional nursing hour per patient day (HPPD), taking into account a systematic approach estimating effort for a patient with a specified disease.
- *Meal order*: The meal order starts with a listing of the in-patients on a ward. Detailed orders according to the needs of the patients can be entered.

2.2 Preliminary enquiry

Before the evaluation study started, a preliminary enquiry was conducted to collect personal data of the potential participants (nurses, doctors, secretaries and other staff of the department). For that purpose a questionnaire was applied addressing computer-experience in general as well as experience with IS-H*MED, area of work, used IS-H*med functions, age and sex. 182 persons completed the questionnaire and were willing to take part in the subsequent evaluation study. Results of the survey were treated confidentially.

By means of a cluster analysis six "user-types" according to the used IS-H*med functions were discriminated (see table 2). Moreover three user categories were distinguished due to the general as well as the IS-H*med specific experience: Novices, intermediate and expert users.

Table 2: Specification of the user types (percentage in brackets illustrate how many persons of a user type use the corresponding functions.)

User types	Used IS-H*MED functions
„Prevailing medical secretaries" (user type 1) N = 39	View of laboratory and diagnostic findings (67%) Creation of discharge letter (54%) Documentation of physical examinations (51%) Order of medical examinations (49%) Other (15%)
„Physicians "(user type 2) N = 41	Documentation of diagnostic findings (100%) Diagnose related grouping (100%) View of laboratory and diagnostic findings (93%) Order of medical examinations (93%) Creation of discharge letter (90%) Documentation of physical examinations (59%)
"Nursing staff I"(user type 3) N = 60	Meal order (100%) Nursing category (95%) Diagnose related grouping (88%) Order of medical examinations (80%) Other (25%) Documentation of physical examinations (22%) View of laboratory and diagnostic findings (12%)
"Prevailing physicians (user type 4) N = 22	Documentation of diagnostic findings (100%) View of laboratory and diagnostic findings (86%) Creation of discharge letter (82%) Order of medical examinations (82%) Documentation of physical examinations (73%) Other (14%)
"Prevailing nursing staff"(user type 5) N = 11	Other (91%) Meal order (82%) Diagnose related grouping (55%) View of laboratory and diagnostic findings (27%) Documentation of physical examinations (27%) Order of medical examinations (18%)
"Nursing staff II"(user type 6) N = 9	Nursing category (100%) Meal order (100%) Order of medical examinations (89%) Documentation of physical examinations (89%) Other (78%) Diagnose related grouping 22%)

2.3 Main inquiry

2.3.1 Participants and procedure

The evaluation study was conducted in January and February 2003. Participation was voluntary, no financial incentives were offered. We received 132 responses (online as well as paper-and-pencil Questionnaires) from the 182 participants who took part in the preliminary study and from additional spontaneous participants.

After the exclusion of questionnaires with too much missing data (s. chapter 2.3.2 Data analysis) 106 responses remained. Mean age of these participants was 38 years (SD = 8,81; range: 24-61 years). 55 persons (51,9 %) were female, 36 (34 %) male; 15 participants (14,1 %) did not answer the question about their gender. According to computer-experience, the sample included 22 novice, 27 intermediate and 30 expert users. 27 persons did not give information about their general computer experience or their experience with the IS-H*MED system.

2.3.2 Data analysis

Questionnaires with more than 20% missing data (more than 15 items not answered) were excluded from further analysis (s. Gediga, Hamborg & Willumeit, 1998). In case of less or equal than 15 omissions, missing values were replaced by the mean scale value ('3') of the items. The same procedure was applied if the answer was 'no opinion'. This procedure was controlled by comparing reliabilities based on the records without missing data with the reliabilities based on the records with replaced missing data. The procedure showed no differences of the reliabilities. Some items of the questionnaire (A1, A8, T12, E8, F1, F7, F14, L1, and L7) are formulated negative. The values of these items were inverted by the transformation $r_i' = 6 - r_i$ for further analysis.

To analyse the equivalence of the paper-and-pencil and the online version of the IsoMetrics questionnaire two matched groups (à N = 29) from a sub sample of all participants were established with regard to the user-type, computer experience, age and sex. The equivalence of both formats was assessed with respect to the scale mean values and reliabilities.(For a detailed description of this analysis, see Hamborg, Vehse, Ollermann & Bludau, 2004).

The reliability of the scales was computed in a next step. After that, the mean values for both questionnaire formats were calculated to assess the ergonomic quality of the application. Moreover the IS-H*med profile was compared with the profiles of two reference systems.

For the ergonomic quality of software systems should be assessed with respect to the context of use (user, task, equipment and environment, see ISO 9241 Part 11, 1998) an analysis of variance with user-type and computer experience (user group, experience, etc.) as independent and the 7 IsoMetrics-Scales as dependent variable was calculated. To identify special differences between the identified user-types, post-hoc tests have been calculated. To get more detailed information about ergonomic shortcomings of the software, ratings of the single IsoMetrics items were analysed at least.

2.4 Results

2.4.1 Equivalence of the online and paper-based questionnaire

Analysis of the scale means revealed no marked differences between the two matched samples using the online respectively the paper-and-pencil version of IsoMetrics^S (table 3).

Table 3: Means of the online- and paper-pencil version

IsoMetrics Scale	Format of the Questionnaire	N	Mean	SD	T	df	sig. (2-sides)
Suitability for the task (15)	Online	29	2.54	.793	-.771	56	.444
	Paper	29	2.70	.841			
Self-descriptiveness (12)	Online	29	2.33	.764	-1.559	56	.125
	Paper	29	2.66	.835			
Controllability (11)	Online	29	2.72	.795	-1.423	56	.160
	Paper	29	3.03	.864			
Conformity with user expectations (8)	Online	29	2.87	.634	-1,534	56	.131
	Paper	29	3.13	.692			
Error tolerance (15)	Online	29	2.61	.586	-.683	56	.497
	Paper	29	2.72	.618			
Suitability for individualisation (6)	online	29	1.94	.763	-.245	48.39	.807
	Paper	29	2.00	1.161			
Suitability for learning (8)	online	29	2.52	.708	-.784	56	.437
	Paper	29	2.70	1.058			

Reliabilities (Cronbach´s Alpha) of the IsoMetrics subscales were checked and proved to be at least satisfactory (table 4). As well as the scale means, the reliabilities of the IsoMetrics version are not different except for the subscale "suitability for individualisation" (table 4). Within the scope of power analysis we checked whether the sensitivity of the tests was good enough to detect substantial mean differences. A half point mean difference between the online and the paper-and-pencil format was taken as the lower bound of our interest. Data analysis revealed that all tests would have been able to detect this difference. Because the empirical data didn´t show any difference larger than 0.5 and no significant results, the profiles of both formats were considered as equal.

Table 4: Analysis of reliabilities (Cronbach's alpha) of the paper-and-pencil and the online version of IsoMetrics

IsoMetrics scale	Paper-pencil N = 29	online N = 29	Z (paper-pencil vs. online)	IS-H*med overall means	overall Rel. N = 106
Suitability for the task	0.910	0.921	-0.25	2.77	.90
Self-descriptiveness	0.917	0.901	0.33	2.68	.90
Controllability	0.873	0.849	0.34	2.97	.86
Conformity with user expectations	0.704	0.708	-0.03	3.06	.71
Error tolerance	0.791	0.780	0.10	2.85	.84
Suitability for individualisation	0.962	0.849	2.60*	2.12	.90
Suitability for learning	0.918	0.817	1.55	2.84	.87

The results concerning the mean values and reliabilities corroborate the assumption that the two formats can be treated as equivalent. Therefore data of the online and the paper-and-pencil version were merged for further analysis.

2.4.2 Rating of the systems ergonomic quality

The following results give an overview of the overall rating of the system. The scale mean values of the ratings range between 2 and 3 (except the scale „ conformity with user expectations", which is slightly above, see table 4).

Accordingly the ergonomic quality of the system as assessed by its users can be considered quite low.

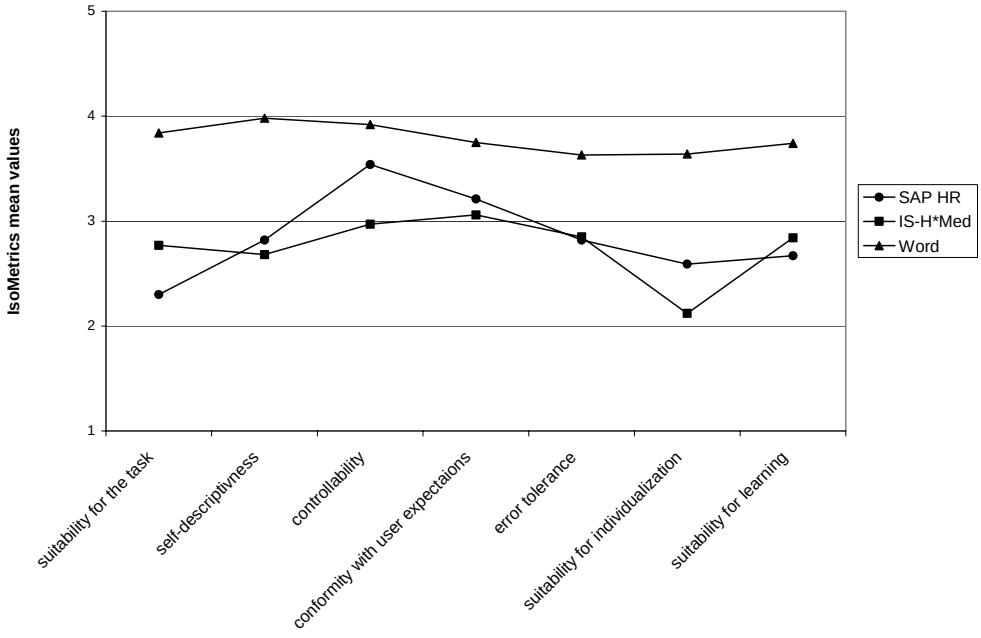

Figure 1: IsoMetrics scale means of IS-H*med and reference systems

2.4.3 Comparison with reference systems

The IS-H*med IsoMetrics mean value profiles were compared with the profiles of two other applications: a) SAP-HR resulting of a study conducted by Gruber (2000) with IsoMetrics, version 2.03 (N = 28) and b) Microsoft Word for Windows (Version 2) which was evaluated in a study by Gediga, Hamborg and Düntsch (1999) with an previous IsoMetrics version (N = 55).

SAP HR is an application supporting several tasks in the field of human resources management like personnel administration, personnel time management, training and event management and payroll accounting. WinWord is the word processing software from Microsoft. The reliabilities of the IsoMetrics scales also were proven to be at least good in this study (table 5).

Table 5: IsoMetrics mean values, standard deviations and reliabilities for SAP R/3 HR (Gruber 2000) and Microsoft WinWord (Gediga, Hamborg & Düntsch, 1999)

	SAP HR			Microsoft Word		
IsoMetrics scale	**Reliability**	**Mean**	**SD**	**Reliability**	**Mean**	**SD**
Suitability for the task	.92	2.30	.72	.53	3.84	0.38
Self-descriptiveness	.84	2.82	.63	.82	3.98	0.36
Controllability	.77	3.54	.50	.79	3.92	0.76
Conformity with user Expectations	.77	3.21	.64	.70	3.75	0.70
Error tolerance	.78	2.82	.52	.63	3.63	0.40
Suitability for individualisation	.73	2.59	.71	.78	3.64	1.02
Suitability for learning	.80	2.67	.67	.54	3.74	0.65

The mean values of IS-H*med and the reference systems were compared using t-values and t-tests respectively. In contrast to SAP-HR, IS-H*med was significantly rated higher on the scale “suitability for the task” and lower on the scales “controllability” and “suitability for individualisation”. It has to be mentioned, that the effect sizes (d, Cohen, 1977) of this tests are quite small. Therefore the observed mean value differences can’t be considered as substantial.

Table 6: Comparison IS-H*med vs SAP-HR

IsoMetrics scale	**Mean**		**t**	**df**	**p**	**d**
	IS-H*med	**SAP-HR**				
Suitability for the task	2.77	2.30	2.96	45.97	0.005*	0.19
Self-descriptiveness	2.68	2.82	-1.04	5.56	0.305	
Controllability	2.97	3.54	-4.62	68.20	< 0.0001*	0.23
Conformity with user expectations	3.06	3.21	-1.02	43.97	0.312	
Error tolerance	2.85	2.82	0.30	54.43	0.762	
Suitability for individualisation	2.12	2.59	-2.85	54.82	0.006*	0.19
Suitability for learning	2.84	2.67	1.13	55.22	0.262	

The comparison of IS-H*med and MS Word reveals significant differences on all scales (empirical values exceed the critical t-value, df = 157, p = 0.007, see table 7). The effect sizes (d) and the mean value differences respectively are large (d > .8, Cohen, 1977, p. 26).

Table 7: Comparison IS-H*med vs MS Word for Windows (Version 2.0)

IsoMetrics scale	**Mean**		**t-Value**	**P**	**D**
	Word	IS-H*med			
Suitability for the task	3.84	2.77	9.39	< 0.0000	1.55
Self-descriptiveness	3.98	2.68	11.48	< 0.0000	1.90
Controllability	3.92	2.97	7.23	< 0.0000	1.19
Conformity with user expectations	3.75	3.06	6.11	< 0.0000	1.01
Error tolerance	3.63	2.85	7.77	< 0.0000	1.28
Suitability for individualisation	3.64	2.12	9.42	< 0.0000	1.56
Suitability for learning	3.74	2.84	6.63	< 0.0000	1.10

It can be concluded, that the usability of IS-H*med is rated low from the user’s point of view and does not meet the ergonomic quality of a standard windows application in this connection but doesn’t differ from other ERP applications substantially.

2.4.4 Evaluation results due to user types

Up to now we have just looked at the pooled ratings of all users taking part in the study. In consideration of the statement that the usability of a software depends on its context of use, we analysed the data with respect to the identified user types.

The IsoMetrics profiles of the different user types are quite similar but located on different levels (figure 2). IS-H*med was rated low especially by doctors while medical secretaries and nursing staff gave the best ratings, but being far from good.

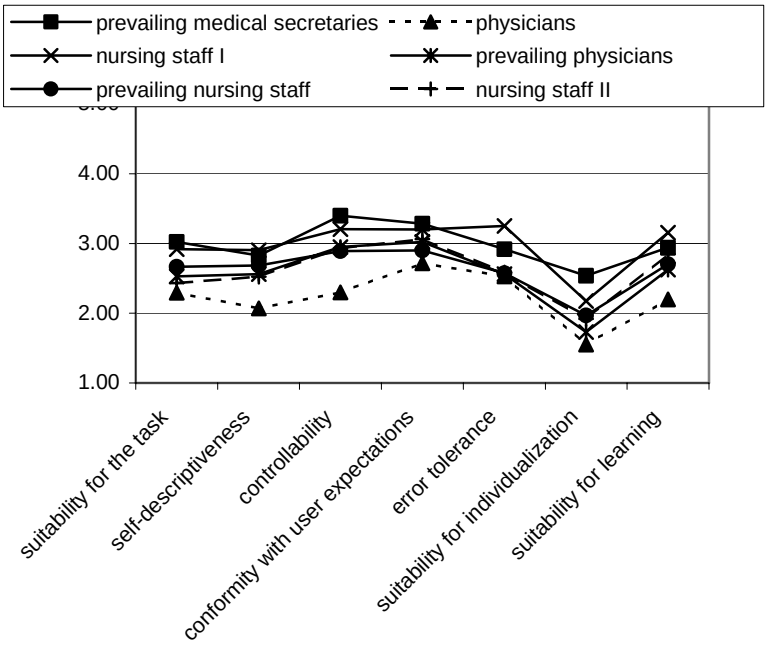

Figure 2: Rating of IS-H*MED according to user types

Statistical analysis shows a significant overall difference between user types on all scales except the scale „Conformity with expectation“. Significant post hoc contrasts were found between „Physicians“ (user group 2) „Medical secretaries” (user group 1), and “Nursing staff I” (user group 3). Physicians rated the systems “Suitability for the task”, “Self-descriptiveness”, “Controllability” and “Suitability for learning” worse than “Nursing Staff I” and “Medical Secretaries”. “Error tolerance” was rated significantly better by “Nursing staff I” (user group 3) compared to “Physicians”.

There are different explanations for this finding. First, by definition of the “user types”, “Physicians” are using other functions than “Medical Secretaries” do. It can be assumed that the functions used by secretaries are of higher ergonomic quality. Especially the MS Word plug-in might have influenced the comparatively good rating of the medical secretaries

Secondly, different ratings might depend on the specific IS-H*med experience of the identified “user-types”. Results of demographic data show that there are more novices than expert IS-H*med users among doctors. In contrast, nursing staff comprises of more experts than novices. A large proportion of nursing staff (43,8 %) has been working with the software for up to 62 months whereas a large part of the doctors (46,8 %) have been using the software for only up to 12 months. The user type of “Medical Secretaries” included no novice users, four intermediate and four expert users. The low experience of the physicians may be due to the fact that the university hospital is a teaching hospital with a high turnover of physicians dependent on a training scheme.

Especially the low ratings of the scales “Suitability for the task”, “Self-descriptiveness”, “Controllability” and “Suitability for learning” received from the physicians should be taken seriously. Its causes should be scrutinised in detail with respect to remedial action.

2.4.5 Item score analysis

So far we have discussed results on the basis of mean value profiles. This kind of analysis allows the assessment of a given software according to standard design principles,

benchmarkings with other applications and the identification of critical usability aspects for specific user groups or task profiles.

In the following we demonstrate an analysis approach that provides some more concrete information about the usability of an evaluated system. It is based on the analysis of the mean values of the single items belonging to the different IsoMetrics scales. This analysis gives a more refined picture of the ratings in comparison to the examination of the scale means presented so far.

To exemplify this approach, we will take a look at the scale “Suitability for the task”. There are some obviously high and low rated items. Analysing the wording of these items in association with the item mean values we get some more detailed information about negative and positive usability aspects of the software.

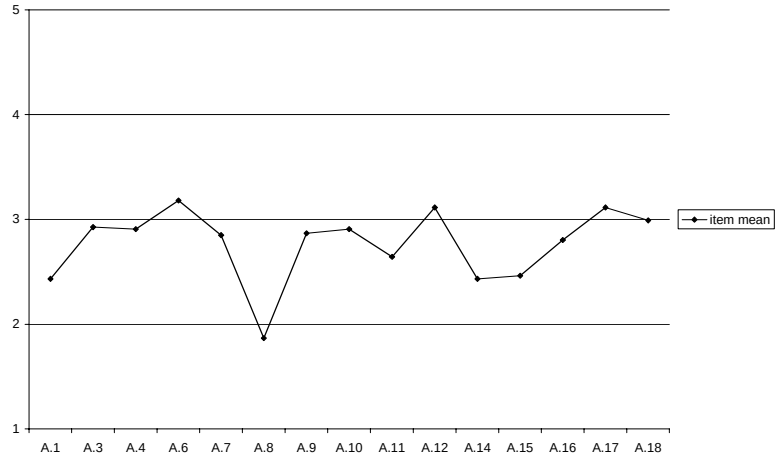

Figure 3: Mean values of the subscale „suitability for the task“

The item means and the standard deviations are shown in brackets after the item description.

Comparatively positive rated items are:

- A.6: The way, in which data is entered is suited to the task I want to perform with the software (M = 3.18; SD = 1.14).
- A.12: The terminology used in the software reflects that of my work environment (M = 3.11; SD = 1.3).
- A.17: I am able to adjust the presentation of results (on the screen, to printer, plotter etc.) to my various work requirements (M = 3.11; SD = 1.26).

Salient negative rated items are:

- A.1: The software forces me to perform tasks, that are not related to my actual work (M = 3.57; SD = 1.28).
- A.8: Too many different steps need to be performed to deal with a given task (M = 4.13; SD = 1.23). (Because the items A.1 and A.8 are formulated negatively the coding has been reversed for statistical data analysis)
- A.14: The software provides me with repeat function for work steps that must be performed several times in succession (M = 2.43; SD = 1.27).
- A.15: I can easily adopt the software for performing new tasks (M = 2.46; SD = 1.12).

Regarding this items, the software appears to conform at least in part to task requirements, especially with respect to data input, adaptability of data output and the wording. Although these items have been rated comparatively positive the system rating was not better than moderate (3). Hence, it cannot be concluded that the software meets the design requirements of this principle (suitability) sufficiently. The low rated items emphasise that the performance of tasks can be awkward and needs extra effort with the evaluated application. Accordingly need for improving the usability of the evaluated software is specified.

3. Discussion

In this paper the IsoMetrics questionnaire was presented. IsoMetrics is a technique to evaluate the usability of software applications with respect to the international standard ISO 9241 Part 10. It could be shown that IsoMetrics is a well suited and reliable technique in the application area of HIS. With help of the online version, that behaves equivalent to the paper-and-pencil format, evaluation studies will become more efficient with respect to the

distribution of questionnaires, data management and data analysis.

IsoMetrics was applied to evaluate a HIS in a field study. Results revealed low ergonomic quality of the evaluated system. The usability of a given software should be treated with respect to given context of use. We demonstrated how to use the IsoMetrics questionnaire to identify critical usability aspects with respect to so called user types, defined by special users dealing with special tasks and functions of the software. Apart from the general low rating of the evaluated software, need for remedial action was identified especially for the group of physicians. The analysis of single items of the questionnaire gives some more concrete hints to weak-points of the evaluated software.

However, principal limits of the application of questionnaires for summative evaluation have to be considered: they will just provide general hints to problem areas of given software but they are not able to detect *concrete* weaknesses nor will they reveal the *causes* why users attribute lack of usability to software. For this purpose, a deeper analysis of problem areas identified by means of a summative questionnaire should be conducted with the help of evaluation techniques like user tests (Dumas & Redish, 1999), walkthroughs (Bias, 1991) or IsoMetrics^{L}.

Especially in large organisations an incremental strategy of usability evaluation seems to be adequate: carry out a screening with a summative questionnaire and identify general problematic usability aspects. After that conduct a more thorough formative evaluation for the identified and most critical usability aspects to reveal the causes for the lack of usability and derive a remedial plan for action.

This kind of systematic evaluation of Hospital Information Systems will support the clinical work of health care employees by adapting the software to user requirements, improving its functionality continuously and avoid errors and stress reactions as well as the costs associated herewith.

4. Acknowledgements

We would like to express special thanks to Alexander Schubek for proofreading this paper.

References

Cohen, J *Statistical power analysis for the behavioural science* (revised edition). Academic Press, New York (1977).

Bias, R `Walkthroughs: Efficient Collaborative Testing´ *IEEE Software,* Vol 8 Issue 5 (1991) pp 94-95.

Dumas, J S and Redish, J C *A Practical Guide to Usability Testing* (Revised Edition), Intellect Books, Exter (1999).

Gediga, G, Hamborg, K-C and Düntsch, I `The IsoMetrics Usability Inventory: An operationalisation of ISO 9241-10´ *Behaviour and Information Technology*, Vol 18 No 3 (1999) pp 151-164.

Gediga, G., Hamborg, K.-C. and Düntsch, I `Evaluation of Software Systems´ In: A. Kent and J. G. Williams, eds. *Encyclopedia of Computer Science and Technology*, Vol 44, Marcel Dekker Incorporated (2002) pp 162-196.

Gediga, G, Hamborg, K-C and Willumeit, H *The IsoMetrics Manual* (Osnabrücker Schriftenreihe Software-Ergonomie Nr. 7). Osnabrück: Universität Osnabrück, Fachbereich Humanwissenschaften (1998).

Gediga, G and Hamborg, K-C `IsoMetrics: Ein Verfahren zur Evaluation von Software nach ISO 9241-10´ In: H. Holling and G. Gediga, eds. *Evaluationsforschung*. Göttingen: Hogrefe (1999) pp 195-234.

Gruber, C (2000). Arbeitseffizienz im Büro. *Psychische Einflüsse auf SAP R/3-unterstützte Arbeitsprozesse*. University of Würzburg, Master thesis [Online]. Available from: http://www.clemens-gruber.de/psychologie/sapr3/ [Accessed: 06.03.2004].

Hamborg, K-C, Vehse, B, Ollermann, F and Bludau, H-B *Softwareevaluation: Zur Äquivalenz der Online- und Papier-Bleistiftversion des IsoMetrics^{S} Verfahrens* (Osnabrücker Schriftenreihe Software-Ergonomie Nr. 8). Osnabrück: Universität Osnabrück, Fachbereich Humanwissenschaften (2004).

ISO 9241-10 Ergonomic requirements for office work with visual display terminals (VDTs): Dialogue principles (1996).

ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs): Guidance on usability specifications andd measures (1998).

Landauer, T K *The Trouble with Computers. Usefulness, Usability and Productivity*. MIT Press, Cambridge MA (1995).

Nielsen, J *Usability Engineering*. AP Professional, Boston (1993).

Developing an Evaluation Instrument for e-Commerce Web Sites from the First-Time Buyer's Viewpoint

Wei-Hsi Hung and Robert J McQueen
Dept. of Management Systems, The University of Waikato, Hamilton, New Zealand
wh9@waikato.ac.nz
bmcqueen@waikato.ac.nz

Abstract: This paper presents the process of developing an evaluation instrument specifically for the evaluation of e-Commerce Web sites from the first-time buyer's viewpoint. The development process is based on theoretical discussions of the Web evaluation and Web user satisfaction literature. A draft evaluation instrument was developed. To enhance its reliability and validity, several iterative trials on e-Commerce Web sites were conducted. Some modifications were made to the instrument. The final version is capable of evaluating e-Commerce Web sites effectively. The instrument provides implications to both Web evaluation practitioners and academics.

Keywords: e-Commerce, Web evaluation, user satisfaction, transaction activity, instrument

1. Introduction

Web-based e-Commerce gives companies global reach and it is far less expensive than alternatives, such as electronic data interchange (Patel et al., 1998). It has become an extremely important avenue for firms in many industries to interact with their stakeholders and customers (Merwe and Bekker, 2003). As the number of transactions through e-Commerce is increasing, the design of Web sites becomes a critical success factor (Kim, Shaw and Schneider, 2003, Wan and Chung, 1998). Forrester Research (cited in Cunliffe, 2000), estimates that poor Web design will result in the loss of 50 percent of potential repeat visits, due to an initial negative experience. Rettig and LaGuardia (1999) suggested that an effective evaluation could lead to better design of electronic systems to meet users' needs. Thus, an evaluation instrument is necessary.

A number of attempts at evaluation of consumer-oriented Web sites have been developed and published in the last few years. Some were in a purely subjective form of individual preferences of the assessor, and some were in the objective form of statistical measurement, such as monitoring the download time of the site and site traffics. However, because Web sites have become more complicated and the number of Web pages has increased, these forms are not able to evaluate Web site effectively. In addition, these evaluation criteria or individual preferences may not be applied to e-Commerce Web sites because this type of Web site requires a consideration of the addition of business-related evaluation criteria (Kim et al., 2003).

This paper presents the process of developing an evaluation instrument specifically for the e-Commerce Web site from a first-time buyer's viewpoint. We define Web site evaluation for the purpose of this paper, as the assessment and measurement of Web sites. Instead of compiling a list of detailed evaluation criteria, this paper chooses crucial criteria based on the discussion on several theoretical models from the business transaction, Web user satisfaction, and Web evaluation literature. The instrument therefore can perform effectively in its evaluation tasks.

This paper starts by reviewing the literature on Web site evaluation, and Web user satisfaction. A Web satisfaction model is then suggested. This paper then discusses how this model can be applied to measure first-time buyers' satisfaction, and how the evaluation criteria and rating systems have been chosen. The methodology used to test the reliability and validity of the model is explained. Finally, an evaluation instrument is presented for the evaluation of e-Commerce Web sites. Its strengths and limitations are also outlined.

2. Overview of web site evaluation

The common issues found in the literature relating to Web site evaluation are quality (e.g. Day, 1997, Loiacono, 1999, Olsina et al., 1999, Rettig and LaGuardia, 1999, Dran et al., 1999, Cox and Dale, 2002, Mich et al., 2003); Web design (e.g. Shneiderman, 1997, Wan and Chung, 1998, Gehrke and Turban, 1999, Thelwall, 2003); and, usability (e.g. Nielsen, 1995, Palmer, 2002, Agarwal and Venkatesh, 2002, Konradt et al., 2003). Researchers have adopted Web quality concept from the quality of product or service (e.g. Loiacono, 1999, Cox and Dale, 2002, Day, 1997). For example,

Dran, Zhang, and Small (1999) adopted Kano's Model of Quality as a theoretical framework to evaluate the quality of Web sites. This model separated product and service quality into three levels according to customer expectations: expected, normal, and exciting. These researchers believe that quality in a product or service is not what the provider or seller put into it, but what the client or customer receives from it. Thus, a Web site should try to satisfy its customers' needs in order to ensure repeat visits from them, and gain their loyalty.

In regard to Web design, Shneiderman (1997) provided an Objects/Actions Interface (OAI) model for Web-site design. This encourages designers of Web sites to focus on analyzing the relationship between task and Web interface. Wan and Chung (1998) looked at problems in Web design from the perspective of network analysis. They suggested that care must be taken when designing the homepage, which is the entrance to the Web site. A homepage should keep the center or median in a Web site. Gehrke and Turban (1999) suggested five major categories that should be considered when designing a Web site for a business: page loading, business content, navigation efficiency, and security and marketing/consumer focus. They argued that page loading is the most important factor in Web-site design. Thelwall (2003) suggested shifting the focus on evaluating Web design from individual pages to aggregated collections based upon Web directories, domains, and entire site.

Undertaking a usability study usually needs high consumer or user involvement, and sometimes the study needs to be conducted in an experimental environment. Nielsen (1993, 1995) provided guidelines and criteria to evaluate the usability of Web sites design and suggested that every design project, including Web site development, should be subjected to usability testing and other validation methods. Toh and Pendse (1997) also suggested that Web pages should be designed for usability and understanding. However, Web sites with good usability cannot guarantee users' preference (Tullis, 1998).

Although some researchers have tried to provide ways of evaluating e-Commerce Web sites specifically (e.g. Boyd, 2002, Merwe and Bekker, 2003), the selection of evaluation criteria still requires more theoretical justification. Overall, frameworks and criteria have been proposed to evaluate e-Commerce Web sites. However, few have evaluated the Web site from a first-time buyer's viewpoint. There is a need to provide theoretical justifications when selecting adequate evaluation criteria for e-Commerce Web sites.

3. Web user satisfaction

For any business, the key to success is repeat business from the same customers (Barnes, 1999). It is the same in the Web environment. A Web site can be considered successful if users are satisfied and revisit it. Satisfied users may spend longer at a Web site, may revisit the Web site later, and may recommend the Web site to others (Zhang et al., 1999). It is crucial to determine what makes a user satisfied with the Web site, as well as what are potential causes of dissatisfaction. To this end, Web evaluators must first know who the users are, what the key goals of those users are, and then they have to know what steps the users are going to take to use that site (Bacheldor, 2000).

3.1 Who are the users?

The users of a Web site are various groups, such as suppliers, buyers, shareholders, or stakeholders. It is also very important to distinguish between first-time, intermittent, and frequent buyers of a Web site (Shneiderman, 1997). For example, first-time buyers usually need an overview to understand the range of services, and to know what is not available, and buttons to select actions. In contrast, frequent buyers demand shortcuts or macros to speed-repeated tasks, compact in-depth information and extensive services to satisfy their varied needs. The user group focused on in this paper is first-time buyers.

3.2 What is first-time buyers' goal?

The goal of first-time buyers is to conduct e-Commerce transaction activities. Several models and frameworks have been proposed to categorize the activities conducted by buyers in the e-Commerce transaction process (Gebauer and Scharl, 1999, Schubert and Selz, 1997, Schubert and Selz, 1999, Lincke, 1998, Liu et al., 1997). For example, Gebauer and Scharl (1999) described the e-Commerce transactions process including information, negotiation, settlement, and after-sales phases. Schubert and Selz (1997, 1999) and Schubert and Dettling (2002) divided online transaction process into information, agreement, settlement, and community phases. Merwe and Bekker (2003) regard the transaction process as consisting of need recognition, gathering information, evaluating

information, and making the purchase. Overall, these models share a certain degree of similarity. This paper has adopted the four-phase model suggested by Gebauer and Scharl (1999). Schubert and Dettling (2002) also adopted this model as a basis to extend their original model. Details of each phase are described as follows.

The information phase comprises both searching for a particular electronic catalog or information, and locating required information and commodities within the Web site. Buyers seek and collect information on potential products or services in this phase. The Web functions supporting the activities in this phase are, for example, the company overview, product catalogs, news releases, and the financial statements.

The negotiation phase serves to establish a contract, fixing details such as product specifications, and payment. Buyers seek transaction information and decision support by assessing the value of special offerings, by identifying new bargaining options, and by increasing the negotiations. The Web functions supporting these activities are, for example, email addresses, phone numbers, fax numbers, and online communication applications that support the buyer to be able to deal with the suppliers online.

In the settlement phase, transaction activities and procedures, which are part of the contract, are comparatively well defined. Web sites to support transaction settlement include extranet systems, and various tools to process orders internally and between transaction partners, facilitate order tracking, and support payment processes. The Web functions are, for example, the payment function, the document exchange, and the order status.

In the after-sale phase, proper access to the transaction file is crucial. Without this, communication problems and delays can occur. The electronic support of after-sale activities is diverse. It ranges from simple electronic mail services to automated helpdesks and sophisticated electronic maintenance manuals. The Web functions supporting the activities in this phase include, for example, the email service, electronic maintenance manuals, FAQs, and training programs.

Based on a hierarchical decomposition (Shneiderman, 1997) of user's activities in these phases, nineteen activities are specified for completing transactions on the e-Commerce Web sites (see Table 1).

Table 1: Activities in each transaction phases

Transaction Phase	Activities
Information	1.find news; 2.find information for specific subjects; 3.find a new product's information; 4.find a new product's price; 5.find a known product's information; 6.find a known product's price; 7.overview company; 8.check financial status.
Negotiation	1.negotiate contract; 2.negotiate price; 3.negotiate volume; 4.negotiate delivery date.
Settlement	1.conduct payment; 2.monitor the goods or services; 3.exchange financial documentation.
After-Sale	1.find maintenance information; 2.ask questions; 3.feedback expression; 4.request training program

3.3 Why are they satisfied?

Kim et al. (2003) suggested that the factors that affect user satisfaction on the Web are attractiveness and informativeness. Attractiveness is defined as the quality of physical settings of the Web site that attracts customers and/or involvement (Kim et al., 2003). It depends on three criteria: customization, interactivity, and vividness. Informativeness is defined as logical settings of the Web, which provide visitors with useful and understandable information (Kim et al., 2003). It comprises three evaluation criteria: understandability, reliability, and relevance.

Zhang et al. (1999) adapted Herzberg's Two-Factor Theory to explain the difference between satisfaction and dissatisfaction. Job dissatisfaction occurs when a group of "hygiene" factors are absent (Zhang et al., 1999). Hygiene factors describe extrinsic factors that impact on employees' relationship to the context or environment where they do their jobs. These hygiene factors remove job dissatisfaction; however, they do not cause people to become highly satisfied and motivated in their work. In contrast, job satisfaction is determined by a group of intrinsic factors named "motivators" (Zhang et al., 1999). Motivators describe employees' relation to what they are doing. One example of this used by Zhang et al. (1999) is that fast loading time will not result in user dissatisfaction, but may not be enough to guarantee user satisfaction.

In addition, Zhang et al. (1999) identified three components contributing to Web user satisfaction or dissatisfaction with a Web interface: information seeking strategy, user characteristics, and Web environment. The strategy or approach a person uses to seek information may be analytic (planned, goal-driven, deterministic, and formal) or browsing (opportunistic, data driven, heuristic, informal and continuous) (Zhang et al., 1999). The Web interface that supports these two strategies is different. For the analytic searching strategy, it is heavily dependent on the functionality of search engine algorithms, while the browsing strategies require a Web user interface that supports "easy and flexible control, high-quality display, and rapid response time". The factors of user characteristic and Web environment can be considered as either a hygiene or motivating factor, depending on the individual differences (Zhang et al., 1999).

4. Development of evaluation instrument

This paper proposes a model to demonstrate and show how an e-Commerce Web site can satisfy its buyers (see Figure 1).

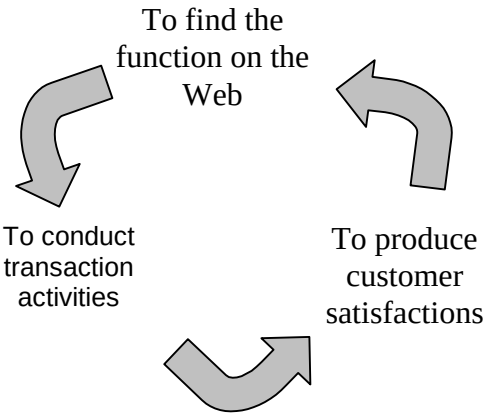

Figure 1: The proposed satisfaction model

This model shows that business buyers will: firstly, find what function or information they want; secondly, use the Web function or information to conduct transaction activities; thirdly, feel satisfied; and finally, find another function or further information. The cycle will continue until buyers finish all their business activities.

Although this model shows how an e-Commerce Web site satisfies its buyers, it does not show how buyers' satisfaction is measured. This paper suggests a two-step process for measuring satisfaction. Firstly, three failure points are identified in the proposed satisfaction model in order to measure whether the Web site can satisfy buyers to complete a transaction activity. Secondly, several evaluation criteria are presented according to the three failure points. They provide more detailed measurements on the degree of satisfaction that buyers perceive when they reach each failure point. The following sections will discuss more details on each step of the two-step process.

4.1 Identify three failure points

Three failure points are identified in the proposed satisfaction model, which are numbered 1, 2, and 3 in Figure 2.

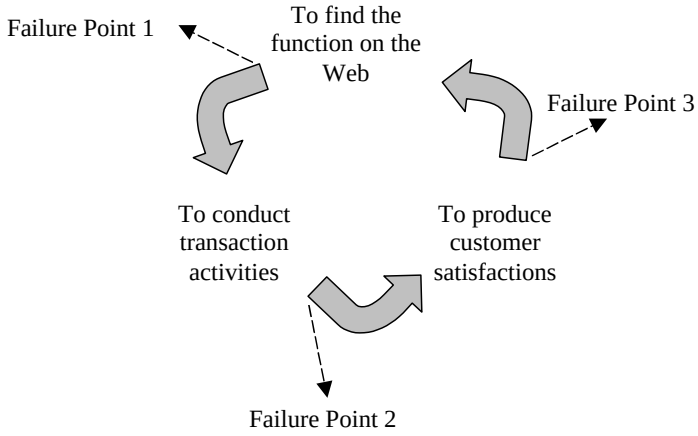

Figure 2: The three failure points in the proposed satisfaction model

Failure point 1 occurs when: buyers cannot access the Web site, or buyers cannot find the function or information they want. Failure point 2 occurs when: the function does not work, or buyers do not know how to use the function or the information is useless. Failure point 3 occurs when buyers do not feel satisfied although they may not be dissatisfied.

All three failure points can be applied to first-time buyers. Because these users are new to the Web site, the failure point 1 can be applied to them when they are using the Web site to perform transaction activities. It is not applied to frequent buyers (see the discussion in Section 3.1) who have conducted some transaction activities before and they know where can find the functions.

These failure points measure different degrees of satisfaction. According to Herzberg's (cited in, Zhang et al., 1999) motivation-hygiene theory, "not dissatisfied" does not equal "satisfied" and "not satisfied" is not the same as "dissatisfied". In other words, there should be a place between satisfied and dissatisfied (see Figure 3).

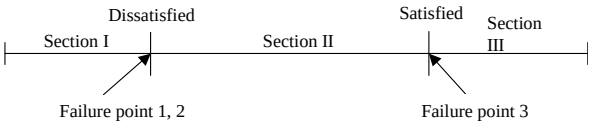

Figure 3: The three failure points and customer satisfaction

Failure points 1 and 2 can be used to measure whether the buyer is dissatisfied with the site. The failure point 3 is used to measure whether the buyer is satisfied with the site. Section I represents those buyers, who are dissatisfied with the Web function, because they can not complete the required activity by using it. Section II covers those buyers who can conduct their activities or the Web function fulfills their needs. However, they do not want to use other functions. Section III represents those buyers who are satisfied by the function. They will try to find other functions to complete the rest of their business activities.

4.2 Choice of evaluation criteria

Although many evaluation criteria are proposed in the literature, the focus here is to choose those criteria which serve to determine whether the first-time buyer can pass through the three failure points in Figure 2 to become satisfied with the Web site. Four suitable criteria have been chosen: ease-of-identification, ease-of-use, usefulness and interactivity. Ease-of-identification is used to measure whether the buyer has passed through failure point 1. Ease-of-use is used to measure whether the buyer has passed through failure point 2. Usefulness and interactivity are used to measure whether the buyer has passed through failure point 3. Table 2 shows the ability of each criterion to measure the three failure points.

Table 2: The ability of each criterion to measure the three failure points

	Ease-of-identification	Ease-of-use	Usefulness	Interactivity
Failure point 1	Yes	No	No	No
Failure point 2	No	Yes	No	No
Failure point 3	Yes	Yes	Yes	Yes

Ease-of-identification has two meanings in this paper. It includes connectivity, and ability of identification. Connectivity is whether the Web site can be accessed reliably and pages load quickly. Ability of identification is the measurement of how easy it is to identify the function from a Web page.

Ease-of-use refers to how easy it is to use the function to achieve the goal of buyers. One of the best ways to illustrate this is to compare moving around in a physical store. For example, buyers are able to get to the checkout counter immediately when they have finished shopping. Similarly, in the Web site, the buyers can always get back to the home page from wherever they are; they also get help quickly when they have questions.

Usefulness refers to whether a Web application would be helpful to the buyers in accomplishing their intended purposes (Lu and Yeung, 1998). Relevant questions are, for example, does it have the functionality which meets buyers' needs; do the Web pages provide sufficient information about the products and services being promoted, such as the size, color, materials, quality?

Interactivity is concerned with how the Web site interacts with the buyers. Three levels of interactivity are identified in this paper: static, dynamic, and interactive contents. Static content, like printed words on a magazine, is a one-way relationship to the buyer (Rachman and Buchanan, 1999). It includes service, and company information. Static content is made only by the Web site provider, and provides the static information, which fulfills buyers' needs. Up-to-date information belongs to this category, for example, new product advertisements, and recent news. Even these types of information change dynamically. However, it is still one-way presentation. Static contents have the lowest interactivity.

Dynamic content is a two-way presentation with buyers. It provides information that instructs or interacts with buyers, for example, customized information and requirement, communication, and transaction functions. Some interactive functions, which include searchable databases, e-mails, and the booking service, are categorized into dynamic content in this paper, because they are a two-way presentation.

Interactive content is a two-way communication between buyers and Web providers in real-time situation. It concerns getting the right information to the right person, in the right format, at the right time. This requires sophisticated Web technology. An example is a chat room, which provides the

communication between the Web provider and its buyers. Functions belonging to this category have the highest level of interactivity.

4.3 Choice of scoring systems

According to the previous discussion on satisfaction and dissatisfaction (see Figure 3), the effectiveness of the four criteria (ease-of-identification, ease-of-use, and usefulness) is based on the degree of satisfaction the buyer perceives. To measure a degree, it is therefore more appropriate to use a multiple-scale scoring system rather than two-scale (e.g. Yes or No). Modified five-point Likert scales have therefore been chosen for this purpose.

Previous work applied original five-point Likert scales (1 to 5) to evaluate the effectiveness of different criteria in Web environment, such as user satisfaction (e.g. Sing, 2004), and ease-of-use (e.g. Misic and Johnson, 1999, Dai and Grundy, 2003, Lii, Lim, and Tseng, 2004). Instead of using the original scoring system (1 to 5), this paper has adapted a system ranging from 0 to 10. The purpose here is to make a larger variance between the results obtained from the Web site evaluation. The evaluator will receive the benefit of it being easier to monitor the gap between superior and poor designed Web functions and Web sites. Together with the criterion interactivity, which is a multiple-choice measure, the scores of each scale for each criterion are shown in Table 3.

Table 3: Scoring systems for the four criteria

Criterion	Levels and Scores				
Ease-of-identification	very easy, right away (10.0)	easy (7.5)	normal (5.0)	very difficult (2.5)	can not find (0)
Ease-of-use	very easy, no help needed (10.0)	easy, no help needed (7.5)	normal, can use, but need help (5.0)	very difficult, need much help (2.5)	do not know how to use or does not work (0)
Usefulness of information	very useful (10.0)	useful (7.5)	normal (5.0)	not useful (2.5)	no usefulness or can not find (0)
Interactivity	interactive content (10.0)	dynamic content (2.0)	static content (2.0)		

Based on the scoring systems in Table 3 and the nineteen transaction activities in Table 1, a draft evaluation instrument was developed. It comprises a series of questions asking how effective the buyer perceived the evaluation criteria after conducting the nineteen transaction activities. Several iterative field tests were conducted to enhance the instrument's reliability and validity. A group of management, computer science, and education students in New Zealand were involved. To enhance reliability, they were asked to evaluate the same e-Commerce Web site within twenty minutes using the draft instrument. If there was a significant difference among their results, some modifications would be made to the instrument. Then, they were asked to evaluate on another Web site until the difference of the results was not significant. To enhance the validity, one management student was asked to use the instrument to evaluate forty e-Commerce Web sites and differentiate them. In the end, many suggestions were received and contributed to modifying the instrument. Some of the nineteen transaction activities identified previously were combined, and fourteen transaction activities were selected. Three instructions were added to the instrument. The final version of the evaluation instrument is shown in Appendix 1.

5. Conclusion

This paper has focused on developing an evaluation instrument for e-Commerce Web sites from a first-time buyer's viewpoint. It has proposed a useful evaluation instrument. As the importance of e-Commerce increases, the instrument will be especially important for those businesses that are currently embracing e-Commerce to evaluate their Web sites. Not only can it differentiate the ability of the site to support first-time buyers to conduct transaction activities, but also measures how well each Web function supports the transaction activities.

The instrument has several strengths. Firstly, it can evaluate different types of e-Commerce Web sites. Guideline-based models are generally grounded on practical experience. These guidelines usually assess "good" or "bad" Web resources, particularly in the usability test. The limitation of this kind of model comes from the difficulty in applying it to various kinds of sites. Compared to this kind of model, the proposed instrument is capable of assessing miscellaneous sites. Secondly, the evaluation instrument does not need to access specific information in the company (such as company's marking strategy) to select evaluation criteria. As Bauer and Scharl (2000) have noted, designing evaluation criteria

usually requires access to company information, which frequently is not available. The evaluation instrument has overcome this difficulty. Thirdly, it is easy to use. Usually, evaluators need specific background about the terms used in the frameworks when using the evaluation frameworks. However, it has been through several iterative tests. Evaluators can use this instrument easily by following the steps and descriptions within it and without knowing specific terms. Fourthly, the evaluation time is less when using the instrument to assess sites in comparison with using other evaluation models (e.g. Merwe and Bekker, 2003). Finally, it is a cheap evaluation instrument in comparison with some evaluation software or services.

However, the proposed instrument has some limitations. Firstly, it assumes that evaluators search information based on a browsing strategy, not an analytical strategy. Buyers with browsing strategy undertake an information seeking approach that depends heavily on the information environment and the buyer's recognition of relevant information. They do not depend on the functions of search engines, unlike the analytical strategy which depends on careful planning, recall of query terms, iterative query reformulation, and an examination of the result (Zhang et al., 1999). Thus, future research should focus on developing another instrument based on an analytical strategy. Secondly, some Web functions may not be accessed because they are password protected or are required to conduct an actual transaction with the company, for example, the order online, chat with a seller, or the payment function. Thus, their usefulness and ease-of-use can not be evaluated fully. Even the proposed evaluation instrument has provided specific criteria to measure. However, the full range of measurement is not created until they are accessed. Finally, platforms of e-Commerce are still in the stage of evolution. Dominant players, such as Cisco, Dell, IBM, and Ariba, are continually developing newer generation of platforms. The fourteen Web functions and transaction activities chosen in the evaluation form might need to be extended in the future. More effective Web functions have to be added in and calculated when one uses the form to measure e-Commerce Web sites.

In conclusion, the evaluation instrument is capable of evaluating e-Commerce Web sites. It is based on a theoretical discussion, and can assist an evaluator to oversee the site easily. This instrument can also be applied to evaluate sites from diverse industries. It can be employed more often to evaluate e-Commerce sites in the future.

References

Agarwal, R. and Venkatesh, V. (2002) "Assessing a firm's Web presence: A heuristic evaluation procedure for the measurement of usability" *Information Systems Research,* Vol 13 No 2 pp 168-186.

Bacheldor, B. (2000) "Web-site design: Simplicity pays for business-to-business sites" *Informationweek,* News Release 14 February 2000, Available from Information.com: http://www.informationweek.com/773/desig2.htm [Accessed 28 August 2000].

Barnes, H. (1999) "Getting past the hype: Internet opportunities for b-to-b markets" *Marketing News,* Vol 33 No 3 pp 11-12.

Bauer, C. and Scharl, A. (2000) "Quantitative evaluation of Web site content and structure" *Internet Research: Electronic Networking Applications and Policy,* Vol 10 No 1 pp 31-43.

Boyd, A. (2002) "The goals, questions, indicators, measures (GQIM) approach to the measurement of customer satisfaction with e-commerce Web sites" *Aslib Proceedings,* Vol 54 No 3 pp 177-187.

Cox, J. and Dale, B. G. (2002) "Key quality factors in Web site design and use: An examination" *International Journal of Quality & Reliability Management,* Vol 19 No 7 pp 862-888.

Cunliffe, D. (2000) "Developing usable Web sites: A review and model" *Internet Research: Electronic Networking Application and Policy,* Vol 10 No 4 pp 295-307.

Dai, X. and Grundy, J. (2003) "Customer perceptions of a thin-client micro-payment system: Issues and experiences" *Journal of End User Computing,* Vol 15 No 4 pp 62-77.

Day, A. (1997) "A model for monitoring Web site effectiveness" *Internet Research: Electronic Networking Applications and Policy,* Vol *7* No 2 pp 1-9.

Dran, G. M., Zhang, P. and Small, R. (1999) "Quality Web sites: an application of the Kano model to Web-site design" *Paper presented at the Fifth Americas Conference on Information Systems*, Milwaukee USA.

Gebauer, J. and Scharl, A. (1999) "Between flexibility and automation: An

evaluation of Web technology from a business processes perspective" *Journal of Computer-Mediated Communication, Vol 5 No* 2, Available from: http://www. ascusc.org/jcmc/vol5 /issue2/gebauer.html [Accessed 7 April 2000].

Gehrke, D. and Turban, E. (1999) "Determinants of successful Web-site design: relative importance and recommendations for effectiveness" *Paper presented at the Proceedings of the 32nd Hawaii International Conference on system Sciences*, Hawaii USA.

Kim, S. E., Shaw, T. and Schneider, H. (2003) "Web site design benchmarking within industry groups" *Internet Research: Electronic Networking Applications and Policy,* Vol 13 No 1 pp 17-26.

Konradt, U., Wandke, H., Balazs, B. and Christophersen, T. (2003) "Usability in online shops: Scale construction, validation and the influence on the buyers' intention and decision" *Behaviour & Information Technology, Vol 22* No 3 pp 165-174.

Lii, Y. S., Lim, H. J. and Tseng, L. P. D. (2004) "The effects of Web operational factors on marketing performance" *Journal of American Academy of Business,* Vol 5 No 1/2 pp 486-494.

Lincke, D. M. (1998) "Evaluating integrated electronic commerce systems" *Electronic Markets,* Vol 8 *No* 1 pp 7-11.

Liu, C., Arnett, K. P., Capella, L. M. and Beatty, R. C. (1997) "Web sites of the Fortune 500 companies: Facing customers through home pages" *Information & Management,* Vol 31 No 6 pp 335-345.

Loiacono, E. T. (1999) "WebQual: a Web quality instrument" *Paper presented at the Fifth Americas Conference on Information Systems*, Milwaukee USA.

Lu, M. T. and Yeung, W. L. (1998) "A framework for effective commercial Web application development" *Internet Research: Electronic Networking Applications and Policy,* Vol 8 No 2 pp 166-173.

Merwe, R. v. d. and Bekker, J. (2003) "A framework and methodology for evaluating e-commerce Web sites" *Internet Research: Electronic Networking Applications and Policy,* Vol *13* No 5 pp 330-341.

Mich, L., Franch, M. and Gaio, L. (2003) "Evaluating and designing Web site quality" *IEEE MultiMedia,* Vol 10 No 1 pp 34-43.

Nielsen, J. (1993) *Usability Engineering*, Academic Press, Boston.

Nielsen, J. (1995). *Multimedia and hypertext: the Internet and beyond*, AP Professional, Boston.

Olsina, L., Godoy, D., Lafuente, G. J. and Rossi, G. (1999) "Specifying quality characteristics and attributes for Web sites" *Paper presented at the International Conference on Software Engineering*, Los Angeles USA.

Palmer, J. W. (2002) "Web site usability, design, and performance metrics" *Information Systems Research,* Vol 13 No 2 pp 151-167.

Patel, J., Schenecker, M., Desai, G. and Levitt, J. (1998) "Tools for growth in e-commerce" *Informationweek,* Vol 712 pp 91-104.

Rachman, Z. M. and Buchanan, J. (1999) "Effective tourism Web sites" *Department of Management Systems Research Report Series*, Number 99-12, University of Waikato, Hamilton, N.Z.

Rettig, J. and LaGuardia, C. (1999) "Beyond 'Beyond Cool': Reviewing Web resources" *Online,* Vol 23 No 4 pp 51-55.

Schubert, P. and Dettling. (2002) "Extended Web assessment method (EWAM): Evaluation of e-commerce applications from the customer's view point" *International Journal of Electronic Commerce,* Vol 7 No 2 pp 51-80.

Schubert, P. and Selz, D. (1997) "Web assessment: A model for the evaluation and the assessment of successful electronic commerce applications" *Electronic Markets,* Vol 7 No 3 pp 1-17.

Schubert, P. and Selz, D. (1999) "Web Assessment: Measuring the effectiveness of electronic commerce sites going beyond traditional marketing paradigms" *Paper presented at the Proceedings of the 32nd Hawaii International Conference on system Sciences*, Hawaii USA.

Shneiderman, B. (1997) "Designing information-abundant Web sites: issues and recommendation" *Human-Computer Studies,* Vol 47 pp 5-29.

Sing, C. K. (2004) "The measurement, analysis, and application of the perceived usability of electronic store" *Singapore Management Reviews,* Vol 26 No 2 pp 49-64.

Thelwall, M. (2003) “A layered approach for investigating the topological structure of communities in the Web” *Journal of Documentation,* Vol 59 No 4 pp 410-429.

Toh, L. Y. and Pendse, S. (1997) “An empirical evaluation of Web page design and consumer behaviour” *Paper presented at the 1st Annual Collected Workshop on Electronic Commerce*, Adelaide Australia.

Tullis, T. S. (1998) “A method for evaluating Web page design concepts” *Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems*, Los Angeles USA.

Wan, H. A. and Chung, C. W. (1998) “Web page design and network analysis” *Internet Research: Electronic Networking Applications and Policy,* Vol 8 No 2 pp 115-122.

Zhang, P., Small, R. V., M, D. G. and Barcellos, S. (1999) “Web sites that satisfy users: A theoretical framework for web user interface design and evaluation” *Paper presented at the Proceedings of the 32nd Hawaii International Conference on system Sciences*, Hawaii USA.

Appendix 1 The Final Version of Web Evaluation Instrument

Step 1: Find all the following Web functions in column 1. If it is found on the homepage, place a "√" in Column 2. If not, jump to the next function.

	Column 1	Column 2	Column 3	Column 4			
No	**Web functions**	**Where?**	**Activity**	**C.1**	**C.2.1**	**C.2.2**	**C.3**
1.1	Company Overview (about us)		To find the information which introduces the company. (then use criteria form 1)				
1.2	Financial Information (investor information or annual report)		To find the financial information about the company. (then use criteria form 1)				
1.3	Privacy (privacy policy)		To find the privacy description. (then use criteria form 1)				
1.4	Product Catalog		To find one product. Is the price shown in the catalog? YES, NO; Can order? YES (jump to 2.1), NO; (then use criteria form 1).				
1.5	New Product Announcement		To find one item of new product. (then use criteria form 1)				
1.6	News (what's new)		To find one item of news. (then use criteria form 1)				
1.7	Learning Information		To find the information which provides knowledge to help learning. (then use criteria form 1)				
2.1	Order (Negotiation)		To find the information about how to order the product. (then use criteria form 2)				
3.1	Payment		To find the information about how to make payment. (then use criteria form 2)				
3.2	Monitoring Goods (order status)		To find the information about how to monitor goods. (then use criteria form 2)				
3.3	Exchange Document		To find the information about how to exchange document. (then use criteria form 2)				
4.1	Maintenance (customer support)		To find the information about how to maintain the product. (then use criteria form 1)				
4.2	Training Information		To find the information about how to train the users of the product. (then use criteria form 1)				
4.3	FAQ of Customer Support		To find the descriptions of FAQ for customer support. (then use criteria form 1)				

Step 2: Conduct those activities, which have ticked, by clicking its function item. Then, complete Column 4 by using criteria forms 1. If other functions are found when conducting activities, write a note to describe under what hyperlink item, into Column 2. If the function is a password protected, then use criteria form 2 to evaluate.

Step 3: Conduct those activities, which are found during the activities, and then complete Column 4 (If the function found does not work, it is scored 0 totally).

Criteria Form 1

Criterion 1: How easy is it to use the function to find one piece of information?	**Criterion 2.2: How useful is the information found?**
A – Very easy. B – Easy. C – Not easy. D – Difficult. E – The function could not work.	A – The content of the information is three times the screen. B – The content of the information is two times the screen. C – The content of the information is one screen. D – The content of the information is less than one screen. E – Useless.

Criterion 2.1: How informative is the Web function?	**Criterion 3: Describe the function and the information found after conducting the activity.**
A – Very informative. The function comprises more than 10 subfunctions. Each subfunction is a hyperlink which links to more specific subjects. B – Informative. The function comprises 5 – 10 subfunctions. Each subfunction is a hyperlink which links to more specific subjects. C – Not very informative. The function comprises 2 – 5 subfunctions. Each subfunction is a hyperlink which links to more specific subjects. D – The function is only a one page presentation. E – Useless.	The function has: A – Search engine: there is a specific search engine provided to search previous information (not the general search function to search the whole Web site) The information found comprises (multiple choice): B – Hyper-links in the text: at least one hyperlinks exists in the final text and provides links to other resources. C – Interactive function: e-mail provided at the end of the information, which is used to inquire about information or give feedback. D – Real-time communication function: there is a function providing communication with the service persons directly.

Criteria Form 2 (for evaluating password-protected functions)

Criterion 1: Is there any helpful instruction provided to guide as to how to use the function?	**Criterion 3: Describe the function and the characteristics found on the Web page where the function is located.**
A – Yes, much helpful information provided, which has more than 10 Web pages. B – Yes, some helpful information provided, which has 2 – 10 Web pages. C – Yes, a little helpful information provided, which has only one Web page. D – Yes, but only the phone number or e-mail address provided. E – No, there is no information which introduces how to use the function.	The function is (choose A or B): A – The function comprises some information, but it does not provide direct interaction with the company. B – It is a function to interact data with company directly. What characteristics are found on the Web page where the function is located? (multiple choice) C – Phone or fax numbers provided at the end of the information, which is used to inquire about further information. D – E-mail provided at the end of the information, which is used to inquire about information or give feedback. E – Real-time communication function: there is a function providing communication with service persons directly.

Scoring Systems for Each Criterion

	Criterion in Criteria Form 1				**Criterion in Criteria Form 2**			
Level	***1***	***2.2***	***2.2***	***3***	***1***	***2.2***	***2.2***	***3***
A	10.0	10.0	10.0	2.0	10.0	–	–	2.0
B	7.5	7.5	7.5	2.0	7.5	–	–	2.0
C	5.0	5.0	5.0	2.0	5.0	–	–	2.0
D	2.5	2.5	2.5	10.0	2.5	–	–	2.0
E	0	0	0	–	0	–	–	10.0

The Evaluation of Software Quality Factors in Very Large Information Systems

Souheil Khaddaj[1] and G Horgan[2]
[1]School of Computing and Information Systems, Kingston University, UK
[2]Nexis Associates Ltd, London, UK
s.khaddaj@kingston.ac.uk.

Abstract: A quality model links together and defines the various software metrics and measurement techniques that an organisation uses which when measured, the approach taken must be sufficiently general for hybrid hardware and software systems. In this work software quality factors that should be taken into account in very large information systems will be considered. Such systems will require a high degree of parallelism and will involve a large number of processing elements. We start by identifying the metrics and measurement approaches that can be used. Many of the quality factors would be applied in similar way for sequential and parallel/distributed architectures, however a number of factors will be investigated which are relevant to the parallel class. In such a system many elements can fail which can have major impact on the system's performance, and therefore it affects the cost/benefit factors. Portability and usability are other major problems that need to be taken into account when considering all the relevant factors that affect quality for such environments.

Keywords: Quality Modeling, Quality Measurement, Software Quality, Very Large Information Systems, Distributed Computing.

1. Introduction

There are a number of requirements that need to be met by a quality model, in order for confidence to be gained that the model correctly captures quality requirements, and correctly reflects how well those requirements have been met. A quality model links together and defines the various software metrics and measurement techniques that an organisation uses. The model answers the question "What is Quality?" and the management of the processes surrounding its use forms a Quality Assurance Process Management Programme, which is defined by the system of policies, procedures and guidelines established by an organization to achieve and maintain quality.

Achieving good performance in a very large information system requires a large amount of computing power, in terms of processors, memory and disk space. It requires the development of techniques for very large scale data handling, efficient strategies for physical clustering of and access to data on secondary storage, and globally distributed data management (Zhang et al, 2003) . This might involves advanced object-oriented modeling and design techniques, and it will require the use of emerging technologies such as computational grids in order to handle such a high degree of complexity (Czajkowski et al, 2001; Von Laszewski et al, 2002).

In the last few decades there have been substantial improvements in computer performance, due not only to advances in hardware but also to innovations in computer architectures, that is, how the computer is designed and organised to carry out its computational tasks. A major development that has affected the computer performance is distributed architectures, wherein the processors are replicated and organised such that they can act in unison on one application. The general acceptance of this technology has been slow for a number of reasons, including the wide differences of the available distributed architectures, which mean that there is no unifying software/hardware environment. It is also widely accepted that software/hardware development can be ad hoc and evolutionary. As a result, engineering environments may start off with poor quality. As the software development evolves so quality should improve.

In this work factors in software quality that should be taken into account when using a very large information system will be considered. We start by identifying the metrics and measurement approaches that can be used. A major example will be the lack of suitable performance metrics, which affect many quality factors such as the cost/benefit factor. Performance metrics for distributed software are tied to the target architecture, and there are as many of these as there are distributed architectures. Portability is another major problem; Changing computer architecture usually requires the rewriting of programs or readapting to a particular architecture and to a particular software environment. Usability is also important since it is harder and more time consuming to program

and use such systems when compared with sequential ones.

2. Quality modelling

Quality is a multidimensional construct reflected in a quality model, where each parameter in the model defines a quality dimension. Many of the early quality models have followed a hierarchical approach in which a set of factors that affect quality are defined, with little scope for expansion (Boehm et al, 1978). More recent models have been developed that follow a 'Define your own' approach (Fenton, 1991). Although an improvement, difficulties arise when comparing quality across projects, due to their tailored nature. The approach used in this work provides a framework in which global and locally defined quality criteria can be considered, individual quality views can be combined and view conflicts can be handled. Thus, it can be used as a standard of excellence measure for a Product, Process or Resource.

Considering that quality comprises implicit and/or explicit attributes, there are a number of views and opinions that need to be considered when defining and measuring quality. These views and opinions are referred to as Essential Views, and are determined by examining an individual's expected use of the product, their experiences in developing or using similar products. By concentrating on removing conflicts of opinion between the Essential Views, a consensus can be reached as to what properties constitute quality, and how quality should be measured.

The properties that constitute the 'explicit and / or implicit attributes' of quality form a set of Key Quality Factors and a set of Locally Defined Factors (Horgan, 2000; Horgan et al, 1999). The Key Quality Factors (KQFs) represent global quality criteria, i.e. factors that are required of all products, and their list of properties is static. The Locally Defined Factors (LDFs) represent local quality criteria, i.e. additional factors identified by the Essential Views, and are appropriate only to the current product being developed. In this way, the KQFs represent a common set of criteria that can be used for cross-project comparisons, whilst the LDFs retain the ability to allow local tailoring (figure 1). The LDFs are not a replacement for the KQFs. Instead, they define additional quality criteria. Their identification and inclusion is entirely the responsibility of the Essential Views. If the different views agree that additional criteria are required, then the additional criteria form the LDFs. In this work we only consider Key Quality Factors and for the rest of the paper they are referred to as simply Quality Factors.

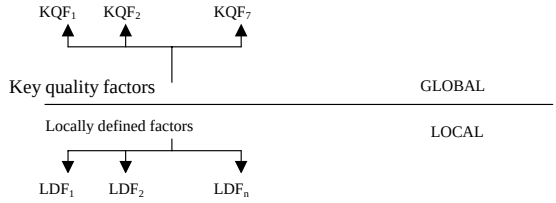

Figure 1: Global and local quality factors

2.1 Quality factors

Quality factors have been used in literature since the early hierarchical quality models (Boehm et al, 1978). The popularity of theses is reflected in the fact that the International Standard ISO 9126 is based on them. The standard recommends a number of factors such Reliability, usability, maintainability etc (Kitchenham, and Plfleeger, 1996). However, people tend to resist plans which evaluate many quality factors, due to limited resources or tight schedules. Based on previous research (Miyoshi and Azuma, 1993) the number of key factors should be kept between three and eight.

In this work a total, eight Quality Factors are defined. These are Performance, Scalability, Cost/Benefit, Usability, Portability, Robustness, Correctness, and Reliability. Many quality factors would be applied in similar way for sequential and distributed systems, as mentioned earlier we are mainly concerned with those specifics to distributed systems. Thus, many factors will not be included such as Maintainability, which is defined as the ability of a product to be modified, and Timeliness, which is defined as the ability of a product to meet delivery deadlines. If a product is delivered late, then it may have good quality aspects, but customers may rightly consider it to be of lesser quality than products delivered on time (Horgan, 2000).

Parallel/distributed processing offers the possibility of an increase in speed and memory beyond the technological limitations of single-processor systems. The performance of such systems can be varied and complex and users need to be able to understand and correct performance problems. Using distributed techniques it is possible to achieve higher throughput, efficiency, performance, and other advantages. Although parallel execution time, concurrency, scalability, and speed-up have been proposed in the literature as performance

metrics, the use of speed-up dominate the literature. The notion of speed-up was initially used to estimate the performance benefit of a multiprocessor over a uniprocessor (Kuck, 1978). Later, speed-up has been proposed as a metric to estimate the performance of parallel and distributed algorithms (Ghosh and Yu, 1998) and parallel processor architectures (Manwaring et al, 1994). The most common definition of speed-up is the ratio of the execution time of the best known sequential algorithm to the execution time of the parallel algorithm on a given number of concurrent processors. The execution time corresponding to a parallel execution is the longest of the CPU times required by any of the concurrent processors.

A general, commonsense definition of performance is presented in (Ferrari, 1978) wherein he defines performance as an indication of how well a system, already assumed to be correct, works. In general, the performance metric for distributed systems must reflect the program, the architecture, and the implementation strategies for it depends on each one of them. However, while the parallel execution time, speed-up, and efficiency serve as well known performance metrics, the key issue is the identification of the distributing processing overheads which sets a limit on the speed-up for a given architecture, problem size, and algorithm.

Every problem contains an upper bound on the number of processors, which can be meaningfully employed in its solution. Additional processors beyond this number will not improve solution time and can indeed be detrimental. This upper bound provides an idea as to how suitable a problem is for parallel implementation: a measure of its scalability. For scalability, we look for the ability to handle large numbers of processes and large or long-running programs on increasing number of processors.

Cost / Benefit is defined as the ability of a product to satisfy its cost/benefit specification. The Costs and Benefits involved in a product's creation should be a major consideration (Turnball, 1991). If the costs are high, and the benefits of its development are low, then there is little point in developing the product.

Usability is defined as the ability of a product to be used for the purpose chosen. It is a factor that is also considered important in other models (Gillies, 1992; Dromey, 1995). If a product isn't usable, then there is little point in its existence. To be useful, a tool should have adequate documentation and support, and should have an intuitive easy-to-use interface. On-line help is also helpful for usability. Adequate functionality should be provided to accomplish the intended task without putting undue burden on the user to carry out low-level tasks.

Because of the short lifespan of high performance computing platforms and because many applications are developed and run in a distributed heterogeneous environment, most parallel programmers will work on a number of platforms simultaneously or over time. Programmers are understandably reluctant to learn a new performance tool every time they move to a new platform. Thus, we consider portability to be an important feature.

For robustness, we expected the product to crash infrequently and its features to work correctly. Research tools are not expected to be as robust as commercial tools, but if the tool has been released for public use, considerable effort should still have been invested in debugging it and on error handling.

Correctness is defined as the ability of a product to meet and support its functional objectives. Other models also include this factor (Fenton, 1991). If software doesn't meet its objectives, then it may be reliable and it may be delivered on time, but no one will use it. Reliability is defined as the ability of a product to reproduce its function over a period of time, and is also included in other approaches (Kitchenham, 1987).

No other factors are included in the Quality Factors list. People tend to resist plans, which evaluate many quality factors, due to limited resources or tight schedules. Based on previous research, the number of key factors should be kept between three and eight (Miyoshi and Azuma, 1993). The Quality Factors set were chosen for their obvious importance for the particular system. However, it is accepted that only empirical validation across a large number of projects can determine the completeness of this set.

2.2 Relationship chart

The first step of the conflict removal mechanism is implemented by use of a Relationship Chart (Gillies, 1992). The chart displays graphically the relationships between quality criteria as a first stage towards measuring the criteria, and provides the basis for constraints on what can be achieved. In the

Relationship Chart, each criterion is listed horizontally and vertically. Where one criterion crosses another, the relationship between those criteria is specified. The relationships for the Quality Factors are fixed. Figure 2 shows the Relationship Chart for the discussed earlier Quality Factors.

By considering these relationships, checks can be made as to the feasibility of requirements. For example, users may state that a reliable product is required, that is both scalable and portable. The relationships between Reliability and Portability, and Portability and Scalability are set to Neutral. Therefore, it is acceptable to state a requirement for a reliable product that is also portable. Similarly, the relationship between Portability and Usability is set to Direct so it is also acceptable to state that a product be portable and usable. As a result, it is an acceptable requirement for a product to be reliable, usable and portable. Note that the Relationship Chart is only a tool and it is still necessary to check the detail of what is being asked.

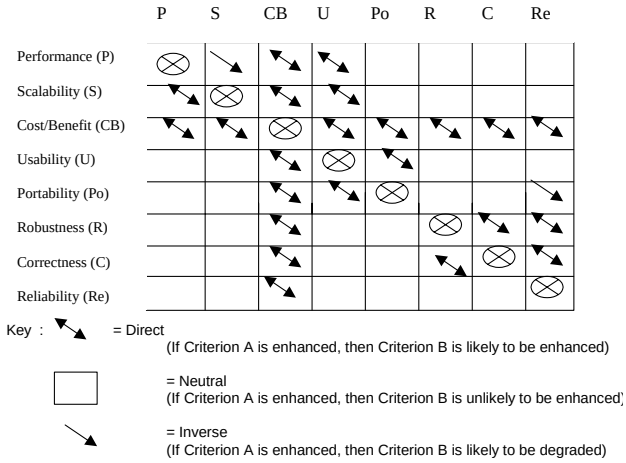

Figure 2: The relationship chart

2.3 Polarity profile

The second step in producing a consensus view of quality is to set the required goals for each criterion, based on the relationships identified in the Relationship Chart. In other approaches, a pie chart is used to represent quality goals (Pfleeger, 1993). There is a need to ensure that anyone can understand the graphical format chosen quickly and easily, particularly when it is considered that some essential views may belong to individuals with little technical background. There is also a need to illustrate over-engineered criteria (i.e., criteria that has exceeded its requirements), since further improvements in these areas will have little effect on the overall quality of the product. Such criteria cannot be shown easily using existing pie chart techniques.

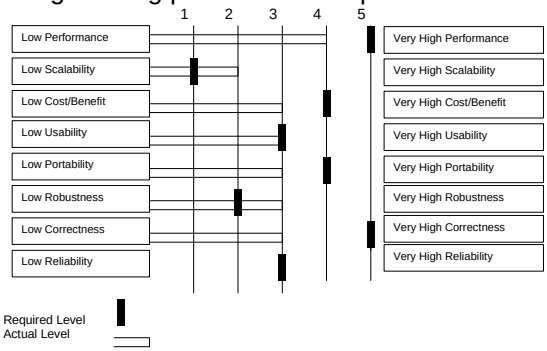

Figure 3: An example polarity profile

The solution chosen, therefore, is to use a Polarity Profile (Gillies, 1992). For each criterion, a range of values exists. The Required Quality of a criterion is defined as a single value on a horizontal line. The Actual Quality achieved is also defined as a single value on the same line. The advantage of using a Polarity Profile is that its format can be easily understood by anyone. Further, it is easy to determine whether or not a criterion has been over-engineered, since its Actual Quality value will be further advanced along the line than its Required Quality value. Figure 3 shows an example Polarity Profile. As can be seen, both Scalability and Robustness have been over-engineered, since their Actual Quality values exceed their Required Quality values. The criteria listed in the Polarity Profile are the same criteria as listed in the Relationship Chart.

Each organisation will use different metrics and metric approaches to measure different quality attributes. These metrics may be similar, identical or entirely different to those used by other organisations. In order to identify the Required Quality for each criterion in the Polarity Profile, the expected properties of that criterion need to be expressed using metrics. The same metrics should be used to identify the Actual Quality for that criterion. There is a need, therefore, for Conversion Mechanisms, which convert the results of metrics used to measure the quality of a criterion. However, for each criterion, the Conversion Mechanism will probably be unique to each metric used. Since different organisations may use different metrics, no single Conversion Mechanism will be suitable in all cases. The Conversion Mechanisms used, therefore, should be agreed between the Essential Views.

3. Conclusions

In the approach presented in this paper individual quality views can be combined, view

conflicts can be removed. It allows the specification of benchmarks against which achieved quality levels can be evaluated, and provides guidance for building quality into software for parallel systems. The feasibility of quality goals is controlled by the use of a Relationship Chart and a Polarity Profile. Since one cannot apply the constraint of having to define upfront the final requirements of a system (Butter, 1998), the approach is not static; if project personnel changes occur, or project requirements change, the Relationship Charts and Polarity Profiles can be updated to reflect these changes. Currently, a set of formal guidelines has yet to be finalised for identifying the Essential Views, despite their importance to the approach. For each occasion that the approach is used, time is required to identify the Essential Views and for those views to derive a consensus of opinions.

References

Boehm, B. W, Brown, J. R, Kaspar, H, Lipow, M, MacLeod, G and Merritt, M. J. (1978) *Characteristics of Software Quality*, North Holland.

Butter, M. (1998) 'Object Disorientation', *Business Computer World*, pp. 107-114.

Czajkowski, K., Fitzgerald, S., Foster, I., and Kesselman C. (2001) 'Grid Information Services for Distributed Resource Sharing' *Proceedings of the Tenth IEEE International Symposium on High-Performance Distributed Computing (HPDC-10),* IEEE Press.

Dromey, R. G. (1995) 'A Model for Software Product Quality', *IEEE Transactions on Software Engineering*, pp. 146-162.

Fenton, N. (1991) *Software Metrics : A Rigorous Approach*, Chapman and Hall.

Ferrari, D. (1978) *Computer Systems Performance Evaluation*, New Jersey: Prentice-Hall Inc., Englewood Clis.

Ferrari, C. (1997) 'The Road to Maturity : Navigating Between Craft and Science', *IEEE Software*, pp. 77-82.

Ghosh, S. and Yu M. L. (1998) 'An Asynchronous Distributed Approach for the Simulation and Verification of Behavior- Level Models on Parallel Processors', *IEEE Transactions on Parallel and Distributed Systems*, pp. 639-652.

Gillies, A. C. (1992) *Software Quality : Theory and Management*, Chapman and Hall.

Horgan, G. (2000) *Construction of a Quality Assurance and Measurement Framework for Software Project*, PhD Thesis, Kingston University, UK.

Horgan, G., Khaddaj, S. and Forte P. (1999) 'An Essential Views Model for Software Quality Assurance', *ESCOM-SCOPE 99*, pp. 387-396.

Kitchenham, B. (1987) 'Towards a constructive quality model', *Software Engineering Journal*, pp. 105-113.

Kitchenham, B and. Plfleeger S. L (1996) 'Software Quality: The Elusive Target', *IEEE Software*, pp. 12-21.

Kuck, D. J. (1978) *The Structure of Computers and Computations*, N.Y., John Wiley &Sons.

Manwaring, M. Chowdhury, M. and Malbasa V. (1994) 'An architecture for parallel interpretation: performance measurements', *Proceedings of the 20th EUROMICRO Conference*, UK, pp 531-537.

Miyoshi, T. and Azuma, M. (1993) 'An Empirical Study of Evaluating Software Development Environment Quality', *IEEE Transactions on Software Engineering*, pp. 425-435.

Pfleeger, S. L. (1993) 'Lessons Learned in Building a Corporate Metrics Program.', *IEEE Software*, pp. 67-74.

Turnball, P. D. (1991) 'Effective Investment in Information Infrastructures', *Information and Software Technology*, pp. 191-199.

Von Laszewski, G., Foster, I., Gawor, J., Schreiber, A and Pena C (2002) 'InfoGram: A Grid Service that Supports Both Information Queries and Job Execution', *Proceedings of the 11th IEEE International Symposium on High-Performance Distributed Computing*, IEEE Press, UK

Zhang, X , Freschl, J. and Schopf, J. (2003) 'A Performance Study of Monitoring and Information Services for Distributed Systems', *Proceedings of HPDC 2003*.

The Impact of IT investment in RSA e-Commerce SME Organisations

Sam Lubbe
Dept. IS&T, University of KwaZulu-Natal, Durban, South Africa
slubbe@ukzn.ac.za

Abstract: This article considers the possibility of a link between organisational performance and information technology (IT) investment intensity in SME organisations practising e-Commerce for the period 2001/2002. The answers to the research questions note that in top performing organisations; (i) IT costs as proportions of operating costs were higher; (ii) IT costs as a proportion of turnover was lower, than in weak performing organisations; and (iii) that a positive correlation exists between the Computerisation Index (CI) and the Operating Costs ratio. The investigation also reveals that Chief Executive Officers (CEO)'s expect additional output while planning e-Commerce operations and keeping IT budgets constant. Evidence is presented that company performance is linked to the level of IT investment intensity in the sample of organisations investigated, even though more output was expected from the IT department.

Keywords: Digital Commerce, e-Commerce, Framework, IT Investment.

1. Introduction

Achieving business value from Information Technology (IT) and e-Commerce investment at the same time is probably one of the more common organisational concerns of (CEOs) today (Lubbe and Pather, 2002). IT and e-Commerce are the growing areas of investment in most organisations; in fact many organisations will not be able to function without IT or digital commerce. The role of IT has also been redefined by some organisations to include attempts to embark on e-Commerce operations. The role of IT in organisations is not merely a tool for processing communication, but a strategic weapon that can thus affect an organisation's competitive position (Weill and Olson, 1989; Lubbe and Pather, 2002).

Some of the variables that will be discussed include IT, e-Commerce, investment and achieving value from IT investment. The contribution of this article is significant as it will contribute to the understanding of managers that the impact of e-Commerce may change the way organisations handle their total IT investment. The article will, however, review only South African organisations and aims to improve on the topic's understanding off IT and digital investment by managers and academics.

2. Review of past research

Mason *et al.* (1997) argue that Information Systems (IS)[1] as a discipline has not yet developed a tradition of historical research. This historical analysis by them broadens the understanding of the processes and designs during which IT is introduced into organisations and the forces the shape IT investment uses. They argue that a dominant design for this shape could be manifested in several ways; a new organisational infrastructure, new functionality, new products, new services, new production functions or new cost structures. The problem with historical analysis is to discover why some organisations led their respective industries in the use, design and application of IT, and why other organisations, having spent millions of dollars achieved modest success rates.

Hu and Plant (2001) argue that the promise of increased advantage was the driving force behind large-scale investment in IT since the 1970's. Current debate continues amongst managers and academics with reference to the measurable benefits of IT investment. Return on Investment (ROI) and other performance measures in academic literature, indicates conflicting empirical findings. They also submit that it would be convincing to infer causality if IT investment in the preceding years is significantly correlated to the performance of the organisation in the subsequent year. Hu and Plant (2001) used the Granger causality model with three samples of organisations and discovered that there was no increase in the level of financial performance. Rather, it is the other way round – increased financial performance lead to increased IT investment.

Li and Johnstone (2002) argue that a manager can use the framework within which the appropriateness of using real options theory in strategic IT investment by systematically

[1] Information Technology (IT) and Information Systems (IS) will be used alternatively and for the purpose of this article will be interpreted as meaning the same whilst discussing the investment of IT.

justifies the use of IT. They classify IT costs and provide some insight about the relationship between technology standardisation and IT investment decisions. Research by Lubbe and Pather (2002) also reflects that managers of organisations are concerned whether their organisation is achieving IT and e-Commerce value from their organisation's IT investment (Figure 1).

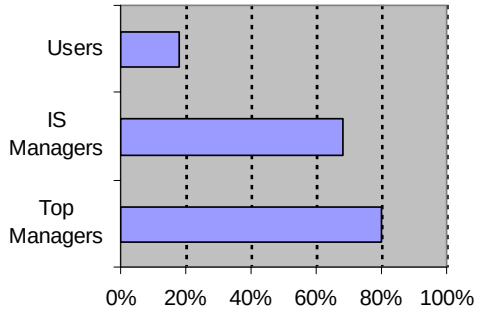

Figure 1: Number of top managers and IS managers concerned about achieving value from IT and e-Commerce

Bui *et al.* (2003) argue that technology and societal changes are moving the global market rapidly towards a new economic order rooted in e-Commerce. They investigate some factors including macro economy, ability to invest, access to skilled workforce, cost of living and pricing. The authors also state that many organisations face a chronic shortage of resources (including funding). Management should be aware that e-Business is part of the complex and general economic structure and the success of organisations depend on that structure as well as the optimum allocation of resources.

Dykman (2003) notes that Information Systems (IS) represents a significant investment for many organisations. Managers need to know that the decision made to spend money on IS should be analysed like any other major purchase. She argues that general management often gives in to the expert power of the technologists, both internal and external to the organisation to invest in IS. The ROI on an IS acquisition may not be quite as simple or straightforward as other capital expenditure. She, however, states that it is still possible to do the financial analysis for the investment.

Dykman (2003) argues that it would be of great benefit if there were a general recipe that could assist to ensure success. Ideally all the strategies (e-Commerce, IT and organisation), including the framing of all investments, could be aligned around business requirements, rather than on technology requirements. She further argues that managers should be measured against the accuracy of their financial projections for IS investment. Every investment should be justified with benefit and expense commitments. A Dykman (2003) note that managers should aim to do good job assessing benefits associated in proposed IS investment in tangible and financial terms. Executives demand this when evaluating the approval, or denial, of any other capital expenditure. IS investment decisions are business decisions and therefore not technology decisions.

Moodley (2003) argues that e-Commerce technologies are becoming increasingly important to South African apparel producers as they are integrated into global value chains. Moodley (2003) suggests that the empirical evidence emanating appears that e-Commerce is still in its infancy but there is potential for growth. The problem is to ensure that there is sufficient financial support to sustain success of e-Commerce. Moodley (2003) argues that South African organisations should increase their investment in e-Commerce.

Quayle (2003) notes that the awareness and level of implementation of e-Business in European Small and Medium Enterprises (SME)'s differ in some aspects from larger organisations. He argues that the issues of highest importance are leadership, time to market, marketing and financial management and a narrow vision of business survival. He further states that small firm's perception of quality, price, production reliability, service reliability and capability to provide support are normal buyer's demands. Nowhere is the aspect of value from IT investments reflected. The idea is that the cost to execute transactions be reduced. He states that developing e-Business expertise is essential to sustain the competitive advantage. SME's must be aware that some aspects such as financial management could impact on their future plans.

It is also argued by Santhanam and Hartono (2003) that the resource-based view can be used to investigate the impact of IT investment on organisational performance. A strong IT capability can support improved organisational performance. Furthermore, their results indicate that organisations with superior IT capability, exhibit current and sustained organisational performance. They note however, that previous performance must be

taken into account while doing these calculations.

Kearns (2004) states that while IT investment has the potential of providing competitive advantage, actual returns on such investment vary widely and a majority of CEO's rank past IT investment disappointing. There are many methods for investment evaluation, but traditional methods do not adequately account for the intangible benefits that characterises strategic investments. They also lack other features of portfolio selection. He describes a model based on the analytic hierarchy process that could possibly overcome the deficiencies associated with traditional approaches to economic evaluation of IT investment. This approach reflects both on tangible and intangible methods and links IT investment to business strategies.

3. The research questions and research methodology

3.1 Research questions

3.1.1 The relationship between profitability (operating expense ratio) and IT investment (IT ratio)

Lubbe and Pather (2002) noted that a relationship exists between profitability and IT expenditures in South African e-Commerce organisations. Quayle (2003) notes that no relationship exists between organisational performance and the relative portion of resources allocated to IT. He argues that the measure of performance will not capture all factors that contribute to the organisation. Using case studies, Weill and Olson (1989) reveal the importance of converting IT investment into productive inputs with different levels of effectiveness, depending on the organisation. There is also empirical evidence that the use of IT results in lower cost (Santhanam and Hartono, 2003). The first research question can thus be formulated as:

> *Is there a negative correlation between IT investment and profitability in e-Commerce intense organisations?*

3.1.2 The relationship between profitability (financial ratios) and Computerisation Index (CI)[2]

Weill and Olson (1989) argue that two key factors are emerging: determining the return on investments (ROI) on IT is difficult; and investment in IT alone is not sufficient. Dykman (2003) suggests that IT investment reduces the cost of revenue generation. IT investment intensity is the level of infiltration of IT into the organisation. Santhanam and Hartono (2003) suggest that evidence indicates that organisational performance is linked to the level of IT investment intensity. This research question specifically compares the overall performance of the organisation with the CI index (another measure of how the organisation computerised their operations) and not the IT Expense (ITEX) ratio as used previously in Question1.

The second research question can thus be stated as (based on the study of Santhanam and Hartono (2003)):

> *Is there a positive correlation between IT investment intensity and organisational profitability?*

3.1.3 The relationship between profitability (return on assets and return on equity) and IT/e-Commerce strategic management integration with organisational strategic management (business management processes)

The third research question is formulated as:

> *Is there a positive correlation between IT investment and strategic management of IT and e-Commerce operations?*

3.2 Research methodology

The author had decided to use qualitative research because it is designed to help him understand the people and the social and cultural contexts within which the organisation operates. To establish the best design it was decided to collect the data needed to answer the research questions discussed above using a structured questionnaire. The population consisted of all IT intensive organisations that have just started an e-Commerce operation during the period 2001/2002. From this list a

[2] Computerisation Index was discussed in detail in a previous paper of Lubbe, Hoard and Parker: *The Profit Impact of IT* (JIT, March, vol. 1, no 10, pp.44 – 51)

number of companies were selected who indicated their willingness to cooperate with the investigation. They were mailed a copy of the questionnaire with a request to include financial statements for the period covered (2001/2002).

The completed questionnaires were analysed to extract the data. The CI was calculated from data collected using the questionnaires. Financial Ratios were calculated using data from the statements and the questionnaires. Data showing the relationships between the CI and measure of financial performance were plotted on graphs using Microsoft Excel.

3.3 Limitations of the study

It is acknowledged that there are other factors could affect the research but the author has decided to limit the study to the papers that were available to him. It was assumed that the organisational financial and other figures, as rendered, were accurate and complete where they could be verified with audit/working papers. Additionally it was assumed that the respondents completing the questionnaire did so accurately. However, a possible source of error lies in the respondents' interpretation of the terminology used in the questionnaire, although it was pre-tested.

Furthermore, some data given by the respondents could not be verified fully, owing to its sensitivity. Also, it was not possible to check the method of accounting and it is acknowledged that this could have influenced some financial ratios. However, given these limitations, it was still possible to use the models to answer the research questions since these sources of error did not differ from those evident in other studies (e.g. Lubbe and Pather (2002) and Weill and Olson (1989)). It was also possible to interpret the results based on the data obtained as no statistical technique could show them to be unreliable.

4. The results

4.1 Information Systems in South Africa in context

South Africa is a medium sized country, 471,000 square miles at the southern tip of the African continent with a population of some 45 million people. Relative to the rest of Africa, South Africa is substantially industrialised. South Africa is a wealthy country from an industrial and agricultural point of view and computers have been actively in use in South African business, education and industry since the early 1960s when both IBM and ICL opened offices in Johannesburg. Today South Africa employs computers in every aspect of industry, business and government as well as having a relatively high percentage of home computers among the middle class. All the major vendors are present and there is considerable interest in hi-tech.

The business and industrial sectors in South Africa are as sophisticated as anywhere in the world in the use of information systems. South Africa leads the world in deep level mining and supports this activity extensively with computer systems. The country also has a substantial financial services sector that has won international recognition for its excellence in information technology. For example the First National Bank (FNB) of South Africa was named one of the world's top 100 computer users by ComputerWorld Magazine in May 1995 and in July 1996 the same bank also won the prestigious Smithsonian Institute prize for the innovative application of biometrics in their Information Technology.

4.2 Discussion of the results

In order to test the validity of aspects of the questionnaire respondents may have had difficulty understanding when answering, a pilot study was conducted using some of the companies in the sample. This was done to ensure that it was possible to collect all data required for the ratios. Ambiguities were removed in order to reflect a concise research instrument.

4.2.1 Research question 1: The relationship between profitability (operating expense ratio) and IT investment (IT ratio)

The data needed for this section was gathered from financial returns provided by the organisations. Figure 2 illustrates a profile of both, the turnover and operating expenses for the organisations in the sample (2001/2002). Turnover exceeds the operating expenses in 2001 as can be seen from Figure 2. However, in 2002, the effects of a low growth rate in South Africa manifests in the turnover slumping to a low. One organisation spent additional resources to expand their operations affecting the overall picture.

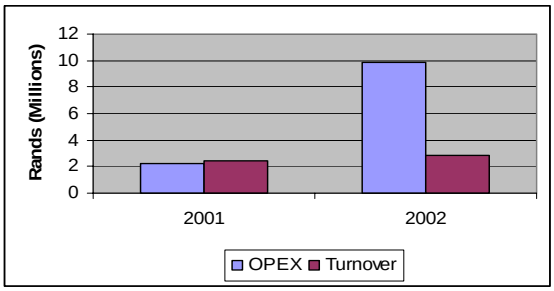

Figure 2: Turnover versus Operating Expense

Operating Expense Ratio (OPEX)[3] and Information Technology expense ratio (ITEX)[4] were the two ratios used in this instance. These were calculated and presented in Table 1.

Table 1: Operating Expense Ratios (OPEX) and IT Expense Ratios (ITEX)

	2001		2002	
Co	**OPEX**	**ITEX**	**OPEX**	**ITEX**
1	0.152	0.119	0.157	0.128
2	0.128	0.037	0.148	0.053
3	0.162	0.117	0.180	0.145
4	0.257	0.160	0.427	0.180
5	0.172	0.483	0.252	0.820
6	0.422	0.139	0.374	0.232
7	0.783	0.118	0.718	0.099
8	0.916	0.002	0.933	0.003
9	0.991	0.002	0.963	0.001
10	0.987	0.003	0.980	0.003
11	1.009	0.062	0.860	0.082
12	0.093	0.001	0.963	0.000

These ratios were calculated and averaged over the period under investigation, to negate the effects of seasonal and abnormal influences as indicated. Finally the organisations were sorted in ascending order using the OPEX ratio as a primary key in and grouped in quartiles (Table 2). This was done partly to disguise the data and to neutralise the effect of seasonal and other influences.

Table 2: Quartile groupings for organisations (2001/2002)

Quartile	OPEX	ITEX
I	0.155	0.100
II	0.317	0.336
III	0.734	0.037
IV	0.965	0.076

As stated before, Table 2 is the result of sorting the organisations (OPEX as the primary key) in ascending order and grouped together in quartiles; the first three companies were used for quartile I, the second three for quartile II, etc. Although all the above-mentioned operations were used to negate the effects of seasonal and economic fluctuations, the results of a loss by one organisation could be seen in the second quartile. There is a negative correlation of 0.5425 between the Operating expense ratio and the IT ratio. This provides evidence that there is a link between the two ratios and supports statements by authors such as Weill and Olson (1989) and Lubbe and Pather (2002).

4.2.2 Research question 2: The relationship between profitability (financial ratios) and Computerisation Index (CI)

Table 3 compares the operating expense ratio, IT expense ratio and CI[5]. The CI indicates and supports the second research question noting that there is a link between computerisation and organisational performance. The better an organisation performs, the higher the CI. From a statistical point of view, the Spearman ranking indicates a high negative correlation of 0.8842 between the CI and the OPEX, while only a positive correlation of 0.4126 was measured between the OPEX and ITEX ratios. CI is therefore a better measure for the intensity of computerisation in an organisation. Lubbe *et al.* (1992) indicated that the CI applies to other industries as well and this further supports this finding.

Table 3: Relationship between CI and Operating and IT ratios

C	CI	OPEX	ITEX
1	73	0.155	0.124
2	47	0.138	0.045
3	13	0.171	0.131
4	10	0.342	0.170
5	7	0.212	0.652
6	6	0.398	0.186
7	5	0.751	0.109
8	5	.0925	0.003
9	5	0.977	0.002
10	5	0.983	0.003
11	8	0.934	0.072
12	6	0.528	0.000

Further statistical analysis indicates an F-Ratio of 3.89 and squared mean deviation of 0.384485 between the CI, OPEX and ITEX ratios. The correlation matrix used to estimate the coefficients produced a correlation-coefficient of -0.8778 between the CI and OPEX ratio and a correlation-coefficient of -0.675 between the CI and ITEX ratio. The correlation was in both instances negative and high. There was also a weak correlation between the CI (the constant, level of computerisation) and the ITEX and OPEX ratios (the variables). It thus helps to answer the second question by delivering proof that

[3] OPEX = non-interest operating expenses to income.
[4] ITEX = IT expenses to non-interest operating expenses.
[5] CI means the extent and sophistication of computerisation. Ten variables (for example years using computers, management activity level, etc.) were selected to collectively represent the computerisation process.

there is a relationship between profitability and computerisation. Figure 3 below illustrates the link between CI and OPEX clearly.

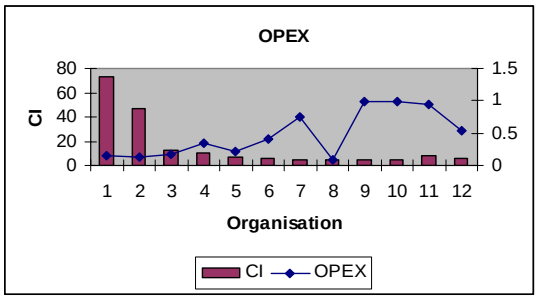

Figure 3: CI versus OPEX ratio

4.2.3 Research question 3: The relationship between Profitability (return on assets and return on equity) and IT/e-Commerce strategic management integration with organisational strategic management (business management processes)

A positive correlation of 0.54 was calculated which led the researcher to accept the fact that there is a relationship between profitability and IT/e-Commerce strategic management integration at the 95% level. A problem that all the respondents mentioned is that they still get the same amount of funding but that top management expects more from them. In real terms, this means that top management expects e-Commerce to stem naturally from the IT department. All the responding organisations placed e-Commerce as part of the IT department.

5. Discussion and conclusion

The relative high correlation that is evident from Figure 3 may be attributed to the strategy employed with IT investment decisions and is supported by Dykman (2003). The strategic importance of IT investment should be emphasised and the importance of IT investment decisions needs to be considered by business managers. The reason being stated is that it may affect their e-Commerce and other commercial operations. Organisations also need to ensure that e-Commerce is not part of the IT department but a department on its own with an own strategy.

It is important to note that the more integrated IT and e-Commerce investment decisions become the better chance for full alignment with the overall organisational strategy. This will help businesses in the long run. Although the study does not conclusively deliver proof of a positive or negative correlation in one instance, it shows that in the sample used, a strong tendency exists that:

- Organisational performance is correlated with IT investment intensity.
- IT investments will be correlated to IT and e-Commerce intensive organisations with their profitability.

It should be noted that to find organisations just embarking on e-Commerce is extremely difficult and explains the reason for the small sample size.

References

Bui T.X., Sankaran S. and Sebastian I.M. 2003: *A framework for measuring national e-readiness*, Int. J. Electronic Business, Vol. 1 No. 1, pp. 3-22

Dykman C.A. 2003: *"Financial Evaluation of Information Systems Investments", Chapter in a Book "Technologies & Methodologies for Evaluating IT in Business" by Charles K. Davis*, Idea Group

Hu Q. and Plant R. 2001: *An empirical study of the Casual Relationship between IT investment and Firm Performance*, Information Resources Management Journal, Jul- Sept pp. 15 - 25

Kearns G.S. 2004: *A Multi-Objective, Multi-Criteria Approach for evaluating IT investments: Results from two case studies*, Information Resources Management Journal, Vol. 17, No. 1, Jan- Mar pp. 37-62

Li X. and Johnstone J.D. 2003: *"Evaluate IT investment Opportunities using Real Options Theory",* IRMJ 15(3), pp. 32-47

Lubbe S., Hoard A. and Parker G. 1992: *The Profit Impact of IT,* Journal of IT, March, Vol. 1, No 10, pp.44 – 51

Lubbe S. and Pather S. 2002: *A Study into Theoretical Success Factors for Successful Internet Commercial Enterprises* – ECITE 2002, held at Université Dauphine, Paris, France, 15-16 July

Mason R.O., McKenney J.L. and Copeland D.G. 1997: Developing a historical tradition in MIS Research, MISQ, September pp. 257 - 277

Moodley S. 2003: *The potential of internet-based business to business electronic commerce for a technology follower: the case of the South African apparel sector*, Int. Journal Electronic Business, Vol. 1, No.1 pp. 75-95

Quayle M. 2003: *e-Business in a turbulent world: usage in European small and medium enterprises*, Int. J. Electronic Business, Vol. 1, () pp. 41-52

Santhanam R. and Hartono E. 2003: *Issues in linking Information Technology Capability to Firm Performance*, MISQ, Vol. 27 No.1, pp. 125-153

Weill P. and Olson M.H. 1989: *Managing Investment in Information Technology: Mini Case Examples and Implications"*, MISQ, March

"A Chronic Wound Healing Information Technology System: Design, Testing, and Evaluation in Clinic"

Antonio Sánchez
School of Computing, University of Glamorgan, Wales
asanchez@glam.ac.uk

Abstract: In the UK, chronic wound healing is an area of specialist clinical medicine that operates within the framework of the National Health Service. It has been the basis for the design, testing and evaluation of a prototype system of information and communication technology (ICT), specifically adapted to the domain. Different wound healing clinics were examined using a combination of 'hard' and 'soft' methods to allow a richer perspective of the activity and gain a deeper understanding of the human activity, its relation to the working information system, the existing information technology (IT), and the potential of a comprehensive IT system to manipulate live data in clinic. Clinicians and administration staff were included in all aspects of the process to enhance the design lifecycle and the understanding of the process. An observe, report, plan and act (ORPA) cycle, based on the dictates of action research, was established to accomplish the design and testing of a system that clinicians were comfortable enough with to consider its use in clinic. Three different strategies were applied to evaluate its use in participating clinics. Cultural historical activity theory was used as the main framework to analyse the activity system, and to interpret the clinicians and the systems performance, as well as their evaluation of the experience. Activity breakdown areas are suggested and reasons for them are considered in the light of wound care workers feedback, and the researcher's observations, notes, and analysis.

Keywords: Electronic data manipulation, clinical ICT, information technology evaluation.

1. Introduction

The introduction of information technology (IT) into clinical medicine is not a new problem (Avgerou 1995). Many specialities have been host to the development of systems of information management based on the potential for superior control of data that information and communication technology (ICT) promises. However, in spite of numerous projects that have been deemed more or less successful by the researchers carrying them out (Littlejohns 2003, Heathfield 1998), a look at clinics in the NHS today does not reveal much IT being used to manage the clinical data, which are so important to the treatment and cure of the patient. The root of this discrepancy could lie in the criteria that are used to assess and evaluate the outcome of these interventions, and the contention that evaluation in general is value bound, and hence conditioned by the views of those conducting the research and the original premises on which it is based (Stone 2001).

The potential of a functioning IT system that is specific to wound healing, a discipline of healthcare and clinical medicine, is a promising area of research and development. ICT is slowly finding its way into the clinic (Simpson 1998, Benson 2002, Benson 2002), and clinicians working in wound care have expressed an active interest in the benefits that technology can bring them and their patients. With the NHS's second strategic plan for the introduction of information technology into healthcare in the UK currently underway (The Department of Health 1998) and set to run from 1998-2005 the climate is right to investigate this in more detail.

This paper considers the process of developing a chronic wound healing information technology system (CWHITS) from the perspective of those actively engaged in wound care. The requirements elicitation, design and testing strategies will draw on a combination of different methods from both systems and social camps, with the main focus of evaluation being drawn from current theories and work in cultural and historical activity theory –CHAT (this will be referred to as 'activity theory' hereafter) (Engeström 1987). Activity theory provides a framework in which to consider the triad of wound carers, the clinical environment in which they work and the 'instruments or tools' that they use to treat patients.

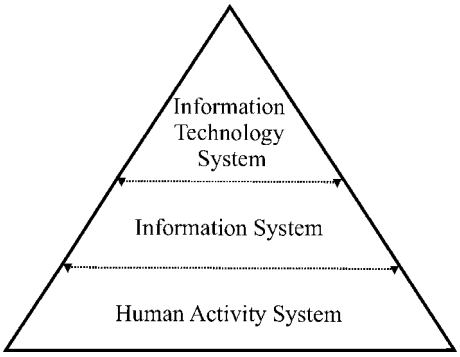

Figure 1: Layers of abstraction of the wound healing activity

In adopting the view of the wound care worker, it is hoped to bridge the gap between the

activity as perceived by the users, the information systems currently in place, and information technology that is overlaid onto that system (figure 1). Devolution of evaluation to the users, and their satisfaction that any system of data manipulation that is put in place helps both them, and their patients, are important considerations for judging any level of success (Thomas 1998). How such a system will accommodate problems, or complications that might arise at the organisational level of the NHS, is a matter that only time and continuing research can resolve.

2. Chronic wounds and consultant clinics in the NHS

Wound healing is probably one of the oldest branches of medicine. Wounds have always been present throughout mankind's existence, and unlike some other medical conditions and problems, they have always been easy to locate and easily assessed in terms of if they have healed or not. The science of wound healing has advanced, however there are still certain wounds that for reasons unknown do not heal as they should. These are termed chronic wounds (Harding 2002) and the savoir-faire to treat and heal them has become the speciality of clinicians who work in chronic wound care. In spite of scientific advances in the drugs and dressings used, as well as to the instruments available, it still remains a complex issue, to fully understand why one wound heals given a certain treatment regime for a particular patient, while another does not.

In the NHS, clinical medicine is the domain of specialist doctor's known as consultants, doctors that have acquired a certain expertise in their particular field, and are regarded as an authority in that area of specialisation. The last 10 years has seen the rise of these specialist clinics, and a shift in the power-base within hospitals and NHS trusts towards these clinics and the consultants that head them (Moss 1995). Unlike administrative healthcare systems, it is generally accepted that clinical specialities do not make full use of the potential of information technology (Benson 2002). The reasons for this are complex and to date inconclusive. This paper expands on existing theories (Huerta-Arribas 1999, Martinko 1996, Silverman 1998, Serafeimidis 2000) and provides new insight into these problems.

3. New technology in the clinic

Clinical healthcare is currently in a state of technological change (Ball 2003). It is only a question of time before technological support tools find their way into the clinic and involve all parts of the medical and healthcare domain. New technology may make this easier, as will user training and technical skill. But to date the numerous documented cases of 'successful' incorporations of technology into healthcare, have been rather limited in scope (Mitchell 2001). What has yet to be seen is a wide scale introduction and implementation of functioning technology tailored to a specific domain in a conclusive way.

Wound healing has already begun to adapt to the incorporation of new computer-based technology. For example, the MAVIS project introduced a tool that allowed non-invasive measurement of wound area and volume using structured light (Plassmann 1998). MAVIS was a device designed specifically for use in wound healing clinics, and in some ways, can be considered as the progenitor of IT in the clinics that it was designed for, and where it was first used. In spite of its main function as a measurement tool, and its limited IT capabilities, this first contact with 'a computer', served to make clinicians, working in those clinics, directly aware of the existence and the potential of computer tools as a benefit to their work activity and their patient's health.

This paper addresses the potential of IT in wound healing, and by prototyping a system that wound care workers feel has taken their needs into consideration, one that they feel can be used in a clinic, to record, access and display, in a reliable structured manner, the different data types that they manipulate, it is hoped that feedback and evaluation will be both insightful and based on values established by those active in the field. Securing the actor's trust (clinicians and administration staff) is vital to achieving successful feedback, and by adopting a policy of inclusion throughout, reporting of the design, testing and evaluation processes, should be clearer and more transparent.

4. Research method

In the case of 'new' information technology, design has tended to centre around the development of new software based on a systems analyst's view of a particular system and user requirements. The tools used for this are based on software engineering precepts, with linear views of the design process. Designers focus on requirement elicitation, software design and testing, systems maintenance and user support. These 'hard' engineering methods lack the scope or depth

of field to include human factors, which are a prime factor in areas of medicine and healthcare.

Nevertheless, these 'hard' methods are necessary if functional software is to be engineered. The real difficulty lies in understanding the systems in place, which the software is to serve. In this work, methodologies and tools, based on understanding the social, political, and organisational aspects of changes to a work activity are used to provide greater insight into the design process. They follow a qualitative research philosophy and have a mainly interpretive view of observed phenomena.

Tools from both 'hard' and 'soft' schools of thought are combined, in the expectation that a richer perspective will give a greater understanding of the relationship between the systems involved (human, information and technology) and produce a richer data set on which to build a cohesive and functioning wound healing information technology system. It should also greatly contribute to any system being considered a viable alternative to paper assessment forms and actually being used by the clinicians.

With regards to practical design, systems thinking and social science methods are now tending to converge on a functional level of application, (McGrath 1998, 2000) but the systems model is more pragmatic and task driven, whereas social methodologies are more concerned with interaction and the process of how things are done, functioning in and as a group, as opposed to as an isolable part of a whole.

5. Strategies of inquiry and sources of data

Depending on the base philosophy they adopt, strategies can fall into two categories (Murphy 1998):

1. Systems theory and software engineering methods provide tried and tested practical tools with which to 'design' an information technology system. In this research entity relationship diagramming, data flow diagramming, and rapid application prototyping have been used.
2. Social science and qualitative research methods provide techniques for immersing the researcher in the social, political, cultural and organisational relationships of a distinct group or system of human activity. In this research ethnography, action research, and activity theory have been applied.

Data was collected at the four participating clinics using a combination of techniques. Depending on which clinic was visited the problem was approached in a different way. The principle sources of data were: Observation: passive and active (semi-participant or involved, and non-interventionist); Interviews (formal and informal); Dedicated focus groups; Questionnaires (structured and semi-structured).

Data was collected, or recorded using a combination of both hard and softcopy tools and media:

Physical (hard)

- Pencil and paper field notes made at wound healing clinics, at focus groups and during interviews with clinicians
- A paper research diary was kept from the start of the study
- Audio tape recordings of focus group sessions and of interviews with wound carers and other personnel
- Completed questionnaires

Digital (soft):

- Data typed on a PDA (personal digital assistant), a Psion V with 8Mb or RAM and a 48Mb smart card was used to enter field notes in clinic, at focus groups, and during interviews and meetings with clinicians and administrative staff
- Recordings were made on the PDA of focus group sessions, interviews, and meetings that took place.
- Voice notes were recorded on the PDA during observation of the clinics
- An electronic diary of the work was kept on a PC and was typed up on the same day as visits to hospitals

6. Participating clinics

The practical implementation of this study took place in four NHS wound healing clinics in England and Wales. They were all led by consultants specialising in the healing of chronic wounds.

Clinic one: a large outpatient clinic based in Wales.

The clinic was managed by the clinical controller, with a PC (personal computer) and access to the hospital trust's PAS (patient admission system), used to keep track of demographic and appointment data. There

were 6 treatment rooms with a communal area where clinicians could consult patient notes, enter observations or dictate their findings to be written up by dedicated 'clinical' secretaries. It was held once a week and was attended by between 30 and 50 patients. Clinical staff consisted of up to 7 wound care nurses, 3 doctors and a consultant, if they were available. The atmosphere was hectic and clinicians had to proceed from one patient to the next without respite. The patient's notes were on a trolley, which had been brought from clinical records by the controller, and were taken in by one of the nurses prior to the patient being summoned.

Clinic two: a medium outpatient clinic based in Wales.

This clinic was also managed by a clinical controller, but with no PC access to the PAS. There were 4 treatment rooms, one being substantially larger than the others, where the clinicians based themselves, to dictate notes or consult patient notes. The clinic was held twice a week and was attended by between 20-30 patients. Clinical staff consisted of up to 6 wound care nurses, 2 doctors and sometimes a consultant. The atmosphere was also hectic, but less so than in clinic one. The patient's notes were in a plastic box with the clinical controller, and the procedure was the same as for clinic one.

Clinic three: a small outpatient clinic based in the west of England.

This clinic took place in only one treatment room; there was no controller and no PC access. Clinical staff consisted of a wound care nurse, an assistant nurse and a consultant. Patient's notes were brought in when the assistant nurse called the patient's appointment. Only one patient was seen at a time, which allowed for the clinicians to dedicate themselves entirely to the patient and their wound.

Clinic four: an inpatient clinic based in the west of England.

This clinic took place on the wards. Treatment was dispensed either at the bedside, or in a nearby treatment room, depending on each individual case and the clinician's assessment of the patient's requirements. As inpatients, sometime treatment could take place up to 3 times in 24 hours, and in each case the consultant would decide how frequently reassessment was required, normally twice a week. Numbers were relatively low, between 7-15 patients were seen by the consultant, and while less intense than clinics one and two, time was an important factor in the clinician's day. The patients notes were collected from clinical records by wound care nurses and stored in 'nurse's rooms' located adjacent to the wards. There were PCs in these rooms, but they were not connected to the hospital's PAS.

7. The wound healing activity: field observation and process modelling

Hospital visits to observe wound healing clinics, and meet clinicians and other NHS personnel took place over a period of 15 months. Clinics were visited regularly, sometimes up to 3 clinics a week, other times none. Focus groups and interviews were organised around staff availability. After initial visits to the clinics and attendance at wound care group meetings, questionnaires were prepared to gage the mood, technical skills, expectations, and have a written record of suggestions from staff. The first set of questionnaires had to be completed again, individually, to correct for clinicians conferring and copying answers, or else providing a collective answer after discussion.

Semi-participant observation of the clinics was undertaken and extensive field notes were recorded, both on paper in a dedicated logbook for each clinic, and digitally recorded on the PDA. A diary was kept from the start. This had an informal structure and was written up immediately after returning from all hospital visits. Additional formal meetings and interviews were carried out with staff responsible for administrative tasks necessary for clinics to function at a hospital level. This included personnel working in clinical records departments and in IT departments. At dedicated focus groups, and after having reviewed the data collected, all clinical parties were gathered together and various strategies were discussed with them as to how to proceed. This involved them in any decisions, and it was hoped would achieve a sense of ownership and inspire use.

Based on the data collected, initial models of wound healing (process, information and data flow) were drawn up. These consisted of entity relationship diagrams and data flow diagrams. These were then shown to wound care workers and explained to them in plain English, to ascertain if they were an accurate representation of their activity. The models

were revised based on feedback from the wound carers and the processes were re-engineered (O'Leary 2000) until a consensus was reached. The models were then synthesised and a compound model for a wound healing information technology system formulated (Sánchez 2004).

The main conclusions reached at the end of this first stage were that any working system would need:

- To have an interface that clinicians were familiar with
- To be able to manage the patient data as structured on the paper wound assessment sheets
- To be able to record dictated voice notes, and ensure their transfer to clinical secretaries
- To index analogue and digital photographs taken of patients' wounds
- To have a system of backing up the data on to a PC, to secure the data, given the limitations of physical memory available on a portable device

8. Introduction of the prototype

After the social and the technical side of wound healing had been expounded and the business process was re-engineered to accommodate the balance between feasible and desirable changes, a practical implementation could be developed, tested, and user feedback and observation used to refine the system. Development of the prototype was an iterative process. It was based on the models produced, and the functional specifications or requirements, as agreed with clinicians in focus groups. This could then be refined in accordance with observations made and feedback received. This cycle of user testing, feedback and observations made, followed by changes to the prototype, adheres to the tenets of action research as prescribed by Baskerville 1998, and with the researchers involved to some extent in all parts of the clinical action, they were able to extensively observe and document the process. This 'ORPA' cycle is represented in figure 2.

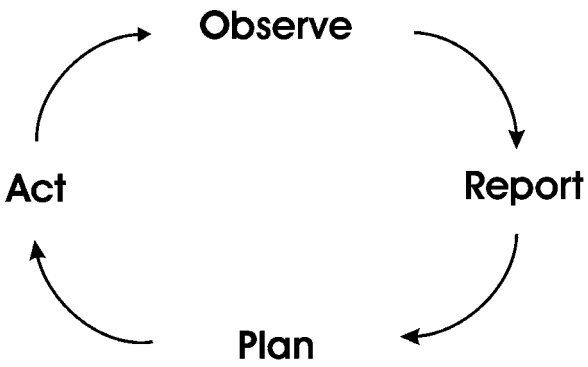

Figure 2: The ORPA cycle used to refine the system

It was agreed that a Psion V MMX with a 96Mb smartcard PDA would be used as the data collection tool. Specific software was written in OPL (organiser programming language), to the agreed specifications, this included a backing-up routine, which would allow for the data collected on the PDA to be transferred via infrared wireless link to a laptop computer, held by each clinic's consultant, and which was to act as a data repository. Ideally this should be done after each clinical session, and at least once every day that it was used.

Once complete the system was tested by the researcher, who found that it could satisfactorily perform the tasks required. User manuals and help files were prepared and the clinicians attended training sessions where they were shown how to use both the PDA and the laptop. Due to the clinicians lack of IT skills, training took longer than anticipated, but at the end they appeared confident and could carry out the tasks necessary to collect data as they did using the paper wound care assessment forms.

The next step was to test the system in live wound healing clinics. There was some concern about this process as clinicians wanted to be sure that it would not compromise the patient's treatment, or that any data be lost. Eventually it was agreed that the data would have to be entered into both the paper assessment forms, and into the PDA. Initial tests involved the researcher entering the data into the PDA, as dictated by the clinicians. The next phase of the test plan was that one clinician would enter the data directly into the PDA, while another entered the same data into the wound assessment sheets. If photos were taken this data was also included. This process was protracted, with the researcher having to answer many questions about usage of the device. The final phase of the test plan was to get the clinicians to use the PDA without any help from the researcher,

who would only be there to observe, and could not intervene. This would be the real test of if IT could be introduced into wound healing clinics in the current climate. The outcome of the experiment is discussed in subsequent sections.

9. Feedback and reporting of the process

The clinician's evaluation on the performance of the CWHITS was injected back into the prototype, which was continually refined through the testing and evaluation process as outlined in figure 3. This process was documented using the same methods as used to observe the activity. The researchers were limited to passive observation for the final tests. The same system of classification was used for the field notes, which were recorded both on the PDA and in the clinical logbooks. An electronic diary was written up for each test run, and this data was invaluable in interpreting the wound care workers actions in terms of their activity. Activity theory was used as a framework for this.

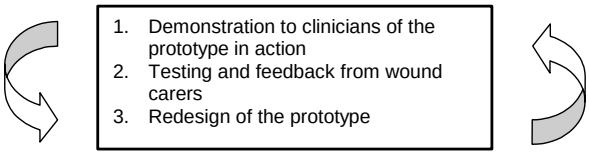

Figure 3: The action research loop

The main ideas offered by clinicians were more conceptual than practical. They were on the whole enthusiastic, but were held back by their limited experience with the technology that was used to replicate existing information systems. Some expressed and interest in attending training courses to obtain basic computer literacy, this was in the form of the ECDL (European computer driving licence). Astonishingly they would have to pay for this themselves.

10. Findings from testing the prototype in clinic

When the prototype was used by the researcher there were no problems of application, as was to be expected. The data was collected electronically on the PDA, and when compared to paper assessment forms the data were the same. When it was a clinician who used the prototype with the researcher's guidance, the data were the same, but the process took much longer, and the clinicians had some problems using the technology. The most important one was being able to see the PDA's screen enough to read it. A solution was found, but it required that the screen be illuminated using the backlight at all times. Battery life thus became a problem, as it was reduced from approximately 18 to 20 hours to only about 2 to 3.

When the clinicians tested the prototype unassisted the result was not positive. They struggled to use the PDA and soon gave up looking for the help files when there was a patient who's wound needed to be dressed. In the end they were unable to balance the limited time that they had to do their job, the need to treat the patient, and their lack of training and IT skills, with the use of this new information system (anecdotally in an attempt to 'make it work' the batteries were removed, and in one instance even the backup battery was removed, which erased all the data stored in memory). This resulted in frustration on behalf of the clinicians and a desire to return to using the assessment forms that they all knew well.

In all three cases there was 'never enough time' to back up the PDA's data, and this resulted in the researcher having to doing it independently. However, even this was not achieved without complications, as the laptop was always 'difficult to get at' and in some cases was even stored in a locked room, which nobody seemed to have the key to!. While a certain persistence ensured that it was eventually done for the first testing strategy, there seemed little point in doing it for the other two, as the researcher was only meant to help the clinicians in the second one, not perform their activity for them, and was meant to be an observer in the third.

As seen there were technical problems of not being able to use the hardware in the clinic as had been envisaged from what had occurred during testing and training. These tended to overshadow the problems of software usage. In the first case there were none, in the second the researcher had to respond to many queries, sometimes repeated, during the examination of a wound, and it was clear that the clinicians were not fully comfortable using the PDA, even when they could see the screen. In the final case, usage of the software was not an issue as, in most cases, the clinicians did not persist in their endeavour to use it, especially with the patient waiting to be attended to.

11. Analysis of the data collected and findings

Analysis of observations made and the data collected has been interpreted using activity theory as a framework, and has resulted in the diagrammatic representation seen in figure 4. This illustrates where the main breakdowns, in this case, secondary contradictions occur in the activity system. The dashed arrows represent problem areas: the relationship between the wound care workers and the data management system, and also between the wound carers and the hierarchical organisation and infrastructure of the NHS. The relationship between the NHS and the clinical information system only has meaning if mediated by the wound carer. As the NHS tries to relate directly to the wound carer they run into problems of contextual definition, which stems from trying to impose something from the top down, instead of trying to combine a bottom up strategy with integration into higher levels of organisational change.

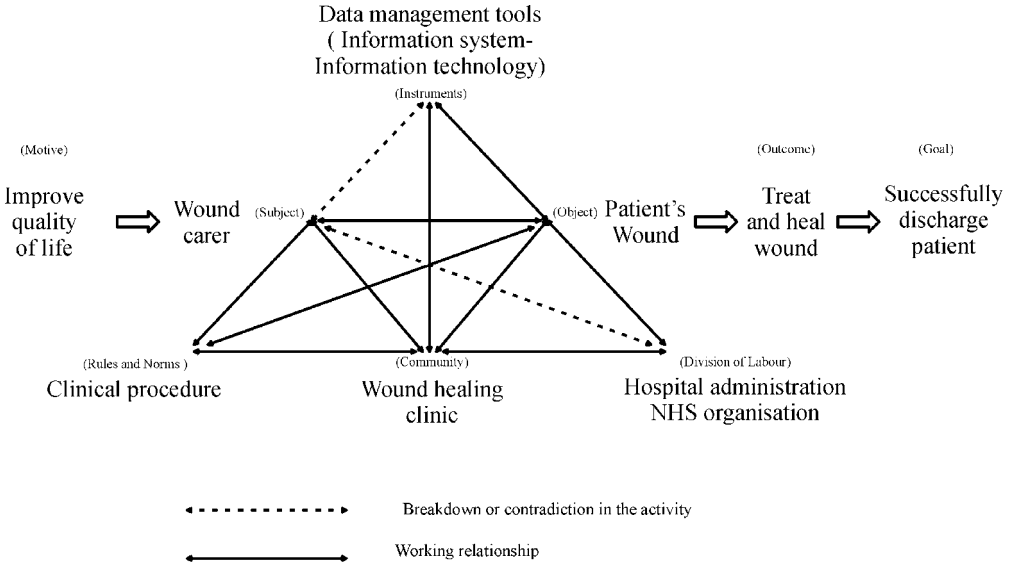

Figure 4: Activity theory model of wound healing in the NHS. (Based on Engeström 1990)

From the point of view of the wound carer, the main contradiction to achieving the object of treating and healing a wound, arises from their relationship with the new instruments, or tools introduced, (CWHITS) and the division of labour inasmuch as wound healing takes place within the structure and organisation of the NHS. The British health minister, Stephen Ladyman (BBC interview 2003), has stressed that the last word with regards to clinical decisions can only be made by the clinician, yet the strategy is already in its fifth and final year, and so far no one has asked the clinicians working in wound healing in hospitals included in this study, what they think about it.

12. Clinicians perceptions of IT in their clinic

In general clinicians seemed keen to think of an IT system as potentially beneficial to their clinical activity. Consultants and doctors were more reserved than nurses, voicing concerns over data security and patient confidentiality (Rindfleisch 1997) - indeed this last point was stressed when obtaining permission to perform this study. They were more practical in their appreciation that there was a gap between their capabilities, existing technology and the technology available. Some believed that any system, not just the one tested was not as reliable as existing hardcopy information systems (even in those clinics where digital cameras are used, photos are printed out, a hardcopy is placed in the notes and the digital photo deleted), and were concerned that 'not having access to their data', could lead to setbacks in the treatment of patients. They failed to realise that they would still have access to the wound data, just that it would be in an electronic format.

Nurses, on the whole, were very enthusiastic about the potential of IT in the clinic, and to their credit, were not dissuaded by their inability to independently use a very simple IT system, the design of which they had participated in. On the whole clinical personal did not appear to feel overly threatened by the new technology, although in one clinic, a nurse had put in place a filing system and was 'in charge' of it, and did perceive the new IT system as a threat to the status quo.

Of the non clinical personal, the only ones who would have direct contact with a new IT system, such as the one proposed here, are the data co-ordinators or clinical controllers, in charge of managing the patient's notes and their appointments, and the clinical secretaries who type up the clinicians' Dictaphone notes. Controllers did appear to perceive the new technology as a direct threat to their jobs, even though they were not involved in the use of the prototype. Clinical secretaries seemed for the most part indifferent, given that the only change to their activity was the media on which the voice notes were recorded. This could change if voice recognition were to be incorporated into the system, and should do so as technological advances take place.

13. Conclusions

In conclusion it is felt that overarching strategies to incorporate technology into clinical medicine can lose sight of their base. By failing to realise that the only ones who can really claim to know what is going on in the clinics are the clinicians, the NHS strategic drive for incorporating IT has failed to take into account that users are the ones that need to be the principle source of consultation. Not the managers, IT designers or other specialists, who have a second hand view of any activity, and whose influence may serve to exclude those who's work it will most affect, and perhaps most important of all, those who's health it could affect.

In this paper, the importance of the conceptual nature of the feedback obtained from clinicians has been stressed. However the current IT strategy for healthcare in the NHS does not seem to include them in its grand designs and instead seems to attribute greater importance to solving managerial or organisational and technical problems related to the desired goal, rather than looking for a realistic and practical solution driven by it.

Lack of proactive IT personal that can motivate clinicians in the hospital trusts where the clinics are based could be one of the main problems in what would be a necessary 'period of transition' from paper to paperless. This accompanied by the apathy that hectic, overworked, understaffed work conditions can induce with regards to "…learning to use a new gadget, when it doesn't help treat the patient or heal the wound, and just takes up more time…" as put by one clinician, does not provide a good foundation on which to build the information strategy for the modern NHS.

14. Final remarks

When undertaking a study of this nature, the researcher felt it important to maintain a certain level of detachment and impartiality, so as to not become attached to any potential prejudices. This was felt necessary, as the researcher was also the designer of the CWHIT, and there was always the potential that their objectivity could be compromised by their desire to see the project succeed - known as 'my baby' syndrome (Littlejohns 2003). The researcher feels that the required level of objectivity was achieved, and that total participation was the only way to give the designer an emic (insiders) view of the world of chronic wound healing in the context of the NHS, and of understanding the somewhat unclear relationship between them.

References

Avgrou, C. (1995) "Evaluating Information Systems by Consultation and Negotiation", International Journal of Information Management, Vol 15, No 6, pp427-436.

Ball, M. (2003) "Enabling Technologies Promise to Revitalize the Role of Nursing in an Era of Patient Safety", Journal of medical informatics, Vol 69, pp29-38.

Baskerville, R (1998) "Diversity in information systems action research methods" European Journal of Information Systems, Vol. 7, pp 90-107.

Benson, T. (2002) "Why general practitioners use computers and hospital doctors do not-part 1:Inventives", British medical journal, Vol 325, pp1086-1089.

Benson, T. (2002) "Why general practitioners use computers and hospital doctors do not-part 2: scalability", British medical journal, Vol 325, pp1090-1093.

Engeström, Y. (1987) "Learning and expanding: and activity theoretical approach to developmental research, Orienta-Konsultit Oy, Helsinki.

Engeström, Y. (1990) "Learning, working and imagining: twelve studies in activity theory, Orienta-Konsultit Oy, Helsinki.

Harding, K. (2002) "Healing chronic wounds", British medical journal, Vol. 324, pp160-163.

Heathfield, H. (1998) "Evaluating information technology in healthcare: barriers and challenges", British medical journal, Vol. 316, pp1959-1961.

Huerta-Arribas, E. (1999) "Evaluation models of information technology in Spanish companies: a cluster analysis",

Information & Management, Vol 36, pp151-164.

Littlejohns, P. (2003) "Evaluating computerised health information systems: hard lessons still to be learnt", British medical journal, Vol 326, pp860-863.

Martinko, M. (1996) "A attributional explanation of individual resistance to the introduction of information technologies in the workplace" Behaviour and information technology, Vol. 15, No. 3, pp313-330.

McGrath, M (1998) "Recording and Analysing Business Processes: An Activity Theory Based Approach", Australian Computer Journal, Vol. 30, No 4, pp 146-152.

McGrath, M. (2000) "Beg, Borrow or Steal: OK but its not all One-Way Traffic!" in ICSE 2000 22nd international conference on software engineering. 2000. Limerick Ireland.

Mitchell, E. (2001) "A descriptive feast but an evaluative famine: systematic review of published articles on primary care computing during 1980-97" British medical journal, Vol. 322, pp279-282.

Moss, F. (1995) "Rethinking Consultants: Alternative models of organisation are needed", British medical journal, Vol. 310, pp925-928.

Murphy, E. (1998) "Qualitative research methods in health technology assessment: a review of the literature", Health technology assessment Vol. 2, No.16 pp57-86.

The NHS Executive (1998) "Information for Health, An information strategy for the modern NHS 1998-2005", [online] Department of health http://www.doh.gov.uk/IPU/strategy/index.htm.

O'Leary, D. (2000) "Knowledge Management For Best Practices" Communications of the ACM, 2000. Vol. 43 No. 11 pp281-292.

Plassmann, P. (1998) "MAVIS - A non-invasive instrument to measure area and volume of wounds", Medical Engineering & Physics, Vol. 20, pp325-331.

Rindfleisch, T (1997) "Privacy information technology and health Care" Communications of the ACM, Vol. 40 No. 8 pp93-100.

Sánchez, A. (1999) "Dynamic medical data fusion and data mining in the wound-healing process", Health Informatics Journal, Vol. 5, No.4, pp213-216.

Sánchez, A (2003) "Evaluation of a mixed media data collection system in clinical medicine" ECITE 2003, Proceedings of the 10th European Conference on Information Technology Evaluation, Instituto de Empresa, Madrid, Spain, 25-26 September 2003.

Sánchez, A (2004) "Practicalities of design, implementation and evaluation of information systems technology specific to chronic wound care clinics in the NHS" PhD thesis, University of Glamorgan, submission pending.

Serafeimidis, V. (2000) "Information systems evaluation in practice: a case study of organizational change" Journal of Information Technology, Vol. 15, pp93-105.

Silverman, D. (1998) "Qualitative research: meanings or practices?" Information systems journal, Vol. 8 pp3-20.

Simpson, K. (1998) "The anatomy of a clinical information system", British medical journal, Vol. 316 pp1655-1659.

Stone, D. (1990) "Assumptions and values in the practice of information systems evaluation", Journal of Information Systems, Fall 1990 pp1-17.

Thomas, R. (1999), "Introduction to Evaluation of Interactive Health Communication Applications", American Journal of Preventive Medicine Vol. 16 No. 1 pp10-15.

