Can Methodological Applications Develop Critical Thinking?

Deborah Blackman¹ and Angela Benson²

¹ School of Management, University of Western Sydney, New South Wales, Australia
² School of Service Management, University of Brighton, Eastbourne, UK

deborah.blackman@canberra.edu.au
A.M.Benson@brighton.ac.uk

Abstract: This paper outlines how using research methods to develop critical thinking was explored in a workshop and then developed into a curriculum. An exercise showed how diverse methodologies led to different answers, which were explored to consider the nature of knowledge itself and the subsequent implications. The paper concludes that such an approach can (a) develop critical thinking skills at a level of deep, rather than surface learning and (b) effectively challenge some preconceived ideas held by students about how knowledge is developed and shared. The crucial element of success was the design and implementation of the assessment.

Keywords: Critical thinking, research methodology, learning, knowledge

1. Introduction

In many universities and schools, an increasing emphasis is being laid upon encouraging more critical and questioning approaches to all studies, which would encourage reflective, deeper learning (Grauerholz, 2001; Reynolds, 1998; James, 1998; Warburton, 2003; Fulop, 2002) and which would develop greater synthesis and analytical skills within the student and/or researcher. However, it has been shown that some management students find the development of such skills challenging (Fulop, 2002). This paper postulates two reasons why critical thinking development is not being successfully implemented into student curricula. The first is that students are encouraged to find the ‘right’ answer (Fulop, 2002), which leads to a rational, often positivistic perspective, which precludes the notion of alternatives. The second, which encourages the first, is that, in order to measure student achievements, assessment seeks to establish if the student has found the “right” answer. This paper firstly, considers ways that may enable alternative answers to be developed and, secondly, considers how assessment needs to be constructed to support criticality. These considerations are then used to show how a curriculum was developed to actively encourage and support critical thinking.

2. How might methodological application develop alternative answers?

When undertaking research, it is usually argued that the framework of the methodology design will be directly relevant to the type and scope of the data collected (Cresswell, 1994; Blaikie, 2000). According to Amaratunga et al. “The overall choice [of methodology] needs, of course, to be the most suitable to achieve the objectives of the specified piece of research” (2002, p.30). As a result of this, it is logical to assume that a different methodology might lead to a different understanding of the same question. Such differences could be explored, compared and contrasted in order to develop a more holistic and considered perspective on complex problems. For this reason, mixed methodologies are increasingly being adopted to expand the answers to research problems (Cresswell, 1994; Denzin, 1970; Denzin and Lincoln, 1994; Onwuegbuzie, 2002). In work using multiple methodologies, the design is usually still driven by the question, rather than using different methodologies to explore whether different answers are discovered to the same problem. However, such approaches do highlight that there will be differences in outputs depending upon the methodologies chosen (Tashakkori and Teddlie, 1998).

Grey and Willmott (2002) argue that the positivist legacy has dominated management research within the United States since the 1950’s and that this has impacted upon the teaching of management worldwide. However, within the social sciences there has been a ‘near collapse of the positivist consensus’ (Grey and Wilmott, 2002, p.2). There is a recognition that there needs to be a growth in the diversity of approaches towards management research and education (Zald, 2002; Grey and Willmott, 2002). The importance of this is stressed by Grey and Wilmott (2002) who argue that the most optimistic positivist must admit that the capacity of management research to produce predictive models has been limited.

With the gradual debunking and discrediting of positivist authority, a space for alternatives has emerged (Grey and Wilmott, 2002). Critical management studies have developed, advancing
ideas to challenge currently accepted perceptions of managerial and business problems. A wide variety of areas have been analysed from a critical perspective, with the core theme being the development of an alternative rhetoric (Parker, 2002). In doing this, researchers ‘borrowed’ from diverse empirical and philosophical disciplines in order to support challenges to ‘the myth of objectivity’, and to argue for a very different, critical, conception of management (Alvesson and Willmott, 1992, pp.3-4). This desire for criticality has permeated throughout much research, publication and into the business and management education curriculum.

However, it is argued that, because management students are often not well enough equipped to be able to undertake such studies effectively, simply presenting students with critical management theory may not lead to the desired outcomes. Mingers (2000) and Fulop (2002) both argue that, in order for there to be a critical approach to management studies, students must achieve the ability to think critically and understand what such skills mean. Without this, student learning goals will be unlikely to move beyond their currently understood model, especially as very often they will have been encouraged to look for one right answer (Fulop, 2002). To change from performance orientated goals to learning goals will take positive encouragement and support (Valle et al., 2003; Ames, 1992). Therefore, not only should critical management studies be within the curriculum but, because critical thinking is rarely something that comes naturally to students, mechanisms for critical thinking teaching need to be developed in a structured way.

In order to teach critical thinking as a skill and to create learning tools which will enable such a skill to be acquired, critical thinking itself will need to be defined (Mingers 2000; Fulop, 2002). Mingers (2000) identifies four dimensions which can be ascertained as encompassing the skills required for critical thinking, in that they enable the questioning of the implicit assumptions or validity claims that should be challenged when applying a critical approach to management:

“First, the logical soundness of the argument and its manner of expression (rhetoric); second, the taken-for-granted assumptions about factual matters and acceptable social practices and values (tradition); third, assumptions made about legitimacy and whose views should be privileged (authority), and fourth, assumptions concerning the validity of knowledge and information (objectivity)” (Mingers, 2000, p. 225).

What will be fundamental, therefore, will be in what ways such questioning can be developed. Various curriculum designs have been outlined (see for example Knights and Wilmott, 1999; Fulop, 2002; Mingers 2000), all encouraging alternative perspectives upon standard management thinking. All discuss epistemology and methodology in order to identify alternative ways of recognising and securing knowledge, but still assume that there should be a choice of methodological application, based upon the epistemological perspective being explored. Zald (2002, p.382), whilst advocating a change in approach to a ‘reflexive/pragmatist epistemology that questions its own grounds’, is still not comparing differences between ideas, rather the choosing of an alternative avenue of research; the application is still being chosen to suit the problem in hand.

However, the question can be raised that, if the desire is to challenge what is currently believed, would it not be possible to do this by demonstrating that different ‘knowledge’ emerges if the same problem is approached from different perspectives. Such recognition would enable several of Mingers (2000) dimensions to be explored at once: the logical soundness would be challenged if different answers were emerging and thus rhetoric would be explored; tradition could be challenged as accepted assumptions could end up in conflict with alternative solutions; issues of authority would emerge, as it would become apparent that different methodologies favoured different stakeholders in a particular situation, whilst assumptions concerning the objectivity and validity of knowledge and information would be challenged. Bhalla et al. (2004) have made some of these arguments when showing how the way that a case study is analysed affects the “knowledge” that develops. The existence of alternative findings and “knowledge” would show that the notion of objectivity needs serious re-evaluation. The difference in this approach from standard triangulation is that, instead of starting with the problem and determining an appropriate methodological strategy, the student would start with a problem and explore what emerges if alternative approaches are applied. The possibility of using methodology in this alternative way offers choices in developing both student learning and, potentially, the possibilities of developing new management research strategies.

3. The role of assessment in critical management education

It is well documented that students are driven by the assessment diet they are presented with.
(Rowntree, 1987; Entwistle et al., 2000), and there is no reason to believe that undertaking critical management studies will be different. Organisational behaviour theory demonstrates that behaviour is triggered from perceived rewards (LeBoeuf 1985), moreover in many of the writings on successful assessment, the focus has been upon learning systems and how assessment will trigger learning (Knight 1995, Brown et al. 1997, Freeman and Lewis 1998). Students respond to what is perceived as important and this, in their perspective, is what is being assessed.

If the objective of undertaking critical management studies is to enable students to perceive the world in a different way, developing alternative mental models of knowledge and potential problem solutions, then the assessment must focus in a way that supports change not confirmation. According to Reynolds “The concept of reflection in experiential learning also embraces an evaluation of alternative explanations and courses of action and of the assumptions on which these are based” (1998, p.3). Reynolds argues that critical reflection is different from other versions of reflection in that “It is concerned with questioning assumptions … the fundamental task of critical reflection is to identify, question and if necessary, change these assumptions” (1998, p. 5). Such critical reflection is described as evolving via techniques, which encourage a range of perspectives. This clearly fits with the notion of changing mental models.

The focus must be upon the process of the assessment and the knowledge development, rather than fixed knowledge outcomes and ‘correct’ answers. If criteria focus upon analysing process as well as outcomes, in order to develop an awareness of the opportunities for alternative forms of “knowledge’ and the possibilities that this implies, the outcomes should support criticality and multiple understandings from the same problem. This paper will now outline the approach that was taken to create a curriculum that would enable critical thinking skills using a methodological approach and using assessment as a primary learning driver.

4. Methodology

There were two phases to the curriculum development. Phase 1 was an exploratory workshop used to identify that the use of methodology gave demonstrably different knowledge outcomes from the same problem. Phase 2 used the workshop outcomes to develop exercises that would enable the development of critical thinking skills using methodology as the framework.

4.1 Phase 1

Initially a workshop entitled ‘How does methodological research inform our organisational debates?’ was undertaken at the British Academy of Management Conference in 2002. The argument being considered was whether the utilisation of alternative methodologies would enable problems to be considered in different ways and, therefore, lead to different solutions being generated, which would, potentially, have implications for organisational study and our understanding of organisational learning. The objectives for the session were given as:

- To explore the role of research methodology in developing new meanings
- To apply such an understanding to improving organisational problem solving
- To explore specific methodologies and consider their contribution to potential organisational learning.

Approximately 45 people undertook an interactive workshop. Groups were asked to identify a specific issue or problem and then apply two or three methodologies, such as ethnography, narrative analysis, metaphor analysis, focus groups, phenomenology, case study development and repertory grids, to that problem. The group identified what was specific about each methodology and then discussed how the use of that approach would facilitate new knowledge and ideas to be used for decision-making. Comparisons between the different outputs enabled the groups to determine whether the application of alternative methodologies would provide useful further study of this approach.

4.2 Phase 2

An exercise was given to approximately 370 second and third year undergraduate students. The disciplines of the students ranged widely, including Business and Management, Marketing, Engineering, Tourism and Computing. The chosen unit explored the relationships between individual learning, organisational learning, knowledge creation and usage and how all of these elements affect, and are affected by, change. A fundamental aspect within the unit was that there must be knowledge developed if there was to be a change in the way people perceived their world and that, for there to be change, new knowledge must be developed. Specifically, the unit outcomes were to:

- Evaluate different conceptualisations of organisational learning;
- Discuss the concept of the learning organisation along with its perceived nature and role;
- Describe processes for knowledge acquisition and management;
- Assess the implications of intellectual capital for management and organization;
- Evaluate theories of learning in relation to knowledge formation and development;
- Synthesise learning and knowledge literatures in order to develop learning and knowledge strategies for organisations and to be able to use this knowledge in order to improve organisational development strategies.

The specific objectives of the exercise were: to consider different knowledge outcomes from different methodologies; to consider different research strategies and the extent to which they would affect learning; to explore how knowledge is developed and shared amongst groups and to reflect on the differences between information and knowledge; to recognise and develop the need for reflection in the effective utilisation of knowledge (Kolb 1984; Dewey 1933).

Each seminar group was split up into six sub-groups with each addressing the question “Why do students choose the course they do at University?” Each group was given a different type of research methodology to approach this question: these were: ethnography; a closed question survey; an open question survey; a case study approach; semi-structured interviews or a literature review. The choice of methodology was based upon the most common methods being undertaken by students in undergraduate dissertations. The groups were given some supporting materials as a starting point and then allocated time to undertake their research. The students were instructed that they would need to
(a) present their findings to their seminar group, (b) compare and discuss their findings with those of other groups in order to discuss the differences and the implications of this exercise upon knowledge development, learning processes and the unit as a whole and (c) prepare for an examination question, which would ask them to reflect upon what they had learnt from the exercise about learning and knowledge and the extent to which this would affect organisational learning and the potential for change.

The exercise was undertaken in class but was then examined in a formal way in order to assess the learning outcomes. By assessing the exercise in the exam, a clear message was being sent that the exercise was important and a fundamental part of the learning for the unit. Such an approach reinforced what had been taught, since in order for learning to take place, which leads to new knowledge creation, there must be a perceived need or desire to learn (Hedburg 1981). In this case, the wish to pass the exam ensured that students would concentrate upon the exercise and think carefully about their learning.

A 5% sample of the exam scripts were analysed in order to determine whether the learning that had been planned had been achieved. The sample was gender balanced and represented students across all marking classifications. Quotes taken from the exam scripts are used as evidence in the following sections.

5. Findings

Outcomes from the workshop can be seen in Table 1. The overall conclusion was that all of Mingers (2000) four dimensions were being challenged in these activities and it appeared that an application of such ideas could be used to enhance critical thinking skills.

Table 1: Workshop outcomes mapped to Minger’s (2000) dimensions

| It is usually considered that the problem should drive the methodology, but the application of other methodologies permits interesting alternative perspectives upon the problem with different apparent knowledge ensuing (challenging objectivity) |
| Each group found that by applying very different methodologies to the same problem, different areas of interest emerged |
| The nature of the problem is seen to be very different when alternative methodological paradigms and methods are used (challenging rhetoric) |
| Diverse assumptions are surfaced when applying different methodologies (challenging tradition) |
| Alternative understandings of the problem and the potential solutions can be elucidated |
| Discussions regarding the use of metaphor analysis for evaluating training effectiveness determined that it would give much more meaningful feedback and enable real analysis |
| It was mooted that the use of alternative methodologies would give different possible futures and this might alter the decisions made |
| Alternative worldviews would be promulgated and this might lead to different perceptions of all aspects of the problem-solving process including the roles of those involved (challenging authority) |

These findings confirmed the proposition that such an approach could be used to develop critical management thinking and a curriculum was developed that would encompass such ideas, using the assessment to support the new focus upon multiple understandings. It was, therefore, posited that successful critical thinking would only
occur if there was deep learning (Warburton 2003) which needed holistic understandings and high motivation, leading to knowledge that would be retained and be applicable to other situations. Consequently, the unit content was mapped onto Warburton’s (2003) model (figure 1) where it could be seen that the inputs should promote the motivation required to achieve the requisite deep learning and change student mental models.

Figure 1: Incentives for deep learning in the critical thinking unit

This curriculum design was then trialled and the findings are presented by examining the four dimensions identified by Mingers (2000). The discussion comes from examining the assessment in order to determine whether the reflection that occurred as a result of the activity, then followed by the assessment, was developing critical ideas in an effective way (Fulop 2002). The second element of analysis was to establish whether the notions of differing knowledge outcomes, as a result of different methodologies effectively challenged student mental models (Hill and Levenhagen 1995) of objectivity and certainty in methodological study.

5.1 Questioning the logical soundness of the argument and its manner of expression (rhetoric)

Students tend to be taught and, therefore, think, in a rather linear way and do not always consider that an argument may be challenged, poorly made or even fallacious. The objective of the exercise was to show that different answers could be gained from varied methodological approaches and alternative methods. By doing this it would be clear that two dissimilar arguments might both appear to be logical but could lead to different, even opposing, solutions. Students recognised the possibilities of differences in several ways. The first realisation was the impact of context and its possible affect upon an argument:

“Although both learning and knowledge are interrelated the outcomes of both concepts may be positive or negative depending on each context they are used in and also the ways in which they are used.”

The role of context was seen as important for both the way the argument worked and in considering how to set up learning and knowledge development:

“In managing both concepts great care must be taken by managers to understand what is meant by O[rganisation] Learning/O[rganisation]K[nowledge] and then contextualise and institutionalise both concepts into the organisation”.

The type of knowledge being developed was also cited as potentially affecting the validity and the certainty of the argument:

“Information varies according to number of people surveyed. That is why other methods such as semi-structured interview, which is a combination of close-ended and open-ended question surveys, which is analysis [of] deep information gained by questioning respondents all critically require proper analysis of information obtained. The explicit knowledge gained is very broad, thus it becomes difficult to cut through it to get to the tacit knowledge involved (Nonaka, 1994). There is no doubt that all methods lead to learning; however, what knowledge is gathered can be questioned. This is because the embedded knowledge may as well have never been found through any of the techniques after all. Looking at explicit knowledge may not lead to any learning after all”.

It was recognised that the explicit knowledge, which was gained by asking respondents the answers, in whatever method, might not reflect any tacit knowledge affecting the answers and this might undermine the validity of the arguments made. Another aspect considered as challenging the possible certainty of an argument was the role of mental models in pre-empting what is learnt (Blackman 2001). An argument can be made that the choice of method is determined, not only by the question to be answered, but also by the previous experiences, beliefs and expectations of the researcher and this was reflected by some students: “what is learnt may not actually be the knowledge attempted to learn because of [existing] mental models which inhibit our learning”: “The observer might be bias[ed] (unintentionally) towards what he is learning due to his mental models”. As a result of the exercise many students recognised that:

“it can be seen that the relationship between [learning and knowledge] is that
the nature of knowledge is relative to the method of learning used”.

In other words the knowledge outcomes gained from the learning processes used depended upon which processes were chosen and how they were implemented.

5.2 Questioning the taken-for-granted assumptions about factual matters and acceptable social practices and values (tradition)

Not a great deal was discussed about this, which was not surprising really from the nature of the exercise given. The discussions about the impacts of the social nature of learning and its impacts could have been linked to this but, during analysis, were more clearly linked to challenging authority or objectivity. What was recognised by students, however, that reflects an outcome in the previous section, was that for there to be effective new knowledge it is necessary to challenge current assumptions and gives:

“Another important relationship I saw is that the level of learning greatly shaped the output. Theorists such as Senge (1990) argued that learning (and thus the knowledge output) is enhanced when mental models are challenged”.

5.3 Questioning the assumptions made about legitimacy and whose views should be privileged (authority)

Several students raised questions about the legitimacy of the knowledge developed due to the fact that different methods lead to different answers and who should decide which of these is privileged:

“it can be seen that each method yields different results on the same things. None of the results are technically ‘wrong’ or the ‘correct’ results. They are just different perspectives on the same thing. As such, one can argue that the nature of the knowledge is relative to the method used”.

This notion of not being able to be sure which results should gain precedence was linked with the issue of observer bias:

“This method [ethnography] is useful to get broad information about certain issues, when information cannot be elicited, as people are either not co-operate or unaware. Although this method also has its limitations, which contain the extent of learning and knowledge gained … It includes a time constraint, as it needs to be longitudinal to have greater accuracy, and also people may change their behaviour when they become aware of the observer, there could be observer bias, you need to be at the right place at the right time, it doesn’t provide the researcher with specific information and may have ethical issues”.

Some students felt that Cook and Brown’s (1999) theory of bridging epistemologies and knowledge development through interactions needed to be considered when looking at the authority of knowledge:

“The use of ethnography may be related to Dewey’s social learning theory, in that through direct and participative observation, ethnographers are interacting in a social context and ultimately learning from their interactions and experiences”.

What they went on to note was that this meant that the researcher and/or observer was an important element of the research outcome as they would affect the choice of learning process, as well as the way that the knowledge was amassed and analysed:

“All methods are open to bias by the researcher”

(The student went on to identify that bias emerges through what is chosen to read, what is asked and how it is interpreted). The process becomes cyclical, affecting and being affected by the experiences undertaken. Thus, it is important to assess the observer and/or researcher as an integral part of the knowledge creation process:

“The learning process itself produces new knowledge and that knowledge impacts upon future learning and thus future possibilities for organisational knowledge”.

Consequently, students outlined the fact that, when managing knowledge creation, the researcher and their biases needed to be assessed and managed as well as other elements of the learning process. Ideas such as groups of researchers, researcher history and a range of methods were proposed in order to reduce the potential biasing in the authority of the knowledge created.

5.4 Questioning the assumptions concerning the validity of knowledge and information (objectivity)

The fourth element that can be challenged, thereby encouraging criticality, is the actual validity and objectivity of the knowledge being developed. This outcome of the exercise using methodology proved to be particularly successful, in that students became very aware of the need to challenge objectivity. The fact that there is no one answer was widely accepted:
“The findings from the exercise proved that there is no one best option or one best way to get the “right” answer. Several areas get linked together: “we can see that each person has their own mental model, perspective and way of learning. They thus each provide a different perspective on an issue. None of these are the “correct” response (indeed Popper (1989) argues no knower can ever truly know) but yet they are not wrong either”.

This shows that the students have really learnt about the potential for challenging the notion of ‘truth’ and a ‘right way’. Students not only assessed the fact that objectivity was debateable but also made suggestions for improvements for future research:

“Thus for organisations to enhance the scope and validity of knowledge used it should use many different methods of learning, i.e. allow more people to provide their perspective and opinions. Indeed, theorists such as Nonaka (1994) argue that this is why team and employee involvement enhance the knowledge output because numerous different perspectives, shaped by each person’s learning method is provided” and “due to the extensive forms of bias in each method it is better to adapt an eclectic approach in using these methods. In order to maximise learning and gather efficient knowledge, the best of each method should be applied and the disadvantages eliminated as efficiently as possible”.

What can be seen examining the Mingers (2000) dimensions above is that students do seem to have learnt a great deal about the nature of knowledge developed via different methodologies. They have recognised that there are different knowledge outcomes from different learning processes and have learnt to challenge the outcomes of apparently valid research.

6. Changing student mental models

Students did develop ideas, which would lead to more critical thinking. However, for this to lead to new patterns of thinking and ongoing challenge in their studies, they needed to have developed new patterns of behaviour and mental models, which would encourage the creation of alternative solutions to similar problems. Evidence of reflection was looked for, which would then lead to a re-evaluation of thinking in a way that would simulate Kolb’s experiential learning cycle (1984) or Argyris and Schon’s double loop learning (1996) and lead to the development of new theories. There was evidence that at least some of the students had developed such reflective practices:

“The exercise] made us realise that different theories and methods and techniques were effective in different ways”; “As a marketing student you tend to see research as more black and white”.

Some students reflected on the problems they saw with what they were doing. Examples include:

“To be in a group of students and just listen to them and observe them while they are discussing and talking about the course chosen is very effective… didn’t have the opportunity to think twice on questions and no one interrupted or led them in their discussion” and “what is learned may not actually represent the knowledge that requires to be grasped. Social learning occurs through experiences and observations, so, all these methods undertaken needs to be adapted in such a way that knowledge is not mistaken or ignored.”

Other students merely indicated that they needed to reconsider things, whilst some began to show what this meant for their future thinking and plans:

“We can increase the scope and validity of knowledge by using many different learning methods. Thus in terms of how they [learning and knowledge] should be managed we should not limit ourselves to one method of learning”.

At the end of the unit it was clear that the students were reassessing the way they thought that knowledge developed. They recognised that in future they would have to develop processes that would permit them to reflect more carefully. The students did not only reflect upon the exercise but also the learning within the unit as a whole. The overall unit reflection led to a picture of learning as an ongoing and changeable process which, depending upon its implementation, would lead to different, probably constructed, knowledge which needs to be regularly challenged.

7. Implications

Table 2: Range of critical factors

<table>
<thead>
<tr>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>The exercise was a summative exercise based upon a body of knowledge about learning and knowledge, which enabled the students to reflect upon the results in the light of their understanding.</td>
<td>A range of both qualitative and quantitative methods was used which forced students to reflect upon different techniques and see how they led to different types of knowledge.</td>
</tr>
<tr>
<td>The unit had been focussed upon challenging student mental models by asking them to read and reflect upon different perspectives on knowledge.</td>
<td></td>
</tr>
</tbody>
</table>
The objective was to test the theory that the use of methodological applications and different methods would enable students to develop criticality; they would be able to learn to challenge assumptions made in arguments and about knowledge. This research indicates that this form of curriculum design would enable such alternative thinking. The success of the exercise was based upon the study of a range of methods and the forcing of real reflection in this exercise. The lecturing team felt that the success of the exercise was owing to a range of critical factors as shown in Table 2.

From this, it can be seen that it was the combination of the mental models developed by knowledge presented in the unit, the methodological content of the exercise, the forced reflectiation via the presentations and the linking of the theory and practice and the fact that the exercise was examined. All these elements came together to lead to a way of developing critical mental models within students. The key is to develop a programme that works on the development of criticality as the desired output throughout the unit. With this learning outcome in mind, it can be concluded that the use of a research methods based exercise is an effective tool that can develop the ability in students to challenge at least three of the assumptions found in arguments and knowledge development. Moreover, student mental models can be led towards more reflective practices and critical thinking over time.

8. Conclusions

This contention of this paper was to outline how to link methodological approaches and critical thinking and to demonstrate the need for further development in the area of teaching critical thinking, especially with the increasing cross disciplinary focus upon teaching management in a more critical way. The development of the assessment tools was fundamental to the success of this undertaking. What is envisaged is that the individuals will become more sensitive to the criticality of methodology and the potential for "knowledge" to be developed in different and, potentially, more interesting ways. The alternative outputs could lead to an awareness that current managerial ways of working may restrict the potential of both individuals and organisations to be innovative, as the linear processes of knowledge development, starting from the problem, prevent a range of ideas occurring. Such awareness may also prove interesting for organisations as well and it is a contention of this paper that, should the approach prove successful with students, it will be an approach that could also be taught to managers in organisations, in order to develop alternative paradigms for innovation development. It is, therefore, hoped that this research will develop a pedagogical approach that can successfully promote critical thinking skills not only in undergraduate students but longer term, within organisations as well. The paper concludes that such an approach can effectively challenge some preconceived ideas held by students about how knowledge is developed and shared. The answers in the exam demonstrated that not only were the different methodologies perceived as developing different outcomes to the same question, but that the implications of this were understood in terms of challenging more traditional approaches to organisational learning and knowledge acquisition.

References


www.ejbrm.com ©Academic Conferences Ltd


Dewey, J. (1933) How we think, DC Heath and Company, US.


Getting the most from NUD*IST/NVivo

Alison Dean and John Sharp
Kent Business School, University of Kent, Canterbury, UK
A.Dean@kent.ac.uk
J.A.Sharp@kent.ac.uk

Abstract: Since the most problematic areas in applying NUD*IST and NVivo in management research are in using them for recording, collating, analysing and reporting interview data, this paper concentrates on the use of NUD*IST and NVivo for these purposes. It examines the problems that arise in their use at the different stages (Interviewing, Transcription, Structuring and Reporting) of three types of research project: a major research project, a specialist research project and a doctoral research project. Suggestions are made as to how these problems can be mitigated. Conclusions are drawn about good practice in the use of NUD*IST and NVivo and suggestions are made on some enhancements that might be made to them. Comments are made about Interpretivism and the use of NUD*IST and NVivo and on the role of supervisors in doctoral research involving the use of NUD*IST and NVivo.

Keywords: Qualitative research reporting; CAQDAS; NUD*IST/NVivo good practice; interview transcription, coding and reporting; interpreting data.

1. Introduction

The focus of this paper is on issues that arise in the application of QSR NUD*IST and NVivo (because of their common descent henceforth NNV) in management research. In recent years, NNV have become popular tools in management research especially amongst young researchers. Objectively, NUD*IST and NVivo were devised to provide a logical way of structuring and enumerating qualitative data. From a computing perspective they are simply a Data Base Management System. Superficially NNV, with their functions for correlating and counting the occurrence of different terms (Richards, T, 2002) might be seen as tools to bring a positivist approach to bear on qualitative data. They support qualitative analysis rather than perform any automated analysis themselves as is sometimes erroneously supposed (Richards, L, 2002).

1.1 Uses of NNV

Computer Assisted Qualitative Data Analysis Software (CAQDAS) is widely used in social science research to facilitate qualitative data analysis. Examples of its use and evaluation are to be found in fields as diverse as marketing research (Doland and Ayland, 2001), sociology (Dohan and Sanchez-Jankowski, 1998; Gibson, Callery, Campbell, Hall and Richards, 2005) and nursing (Morison, 1998; Webb, 1999). Tesch (1991) gives a number of uses of software like NNV in Social Research. These uses are relevant to management research. More specifically, NNV are widely used in management research for the following purposes:
- Organising and analysing literature reviews
- Content analysis of secondary or archival sources

In this paper we concentrate on the use of NNV in relation to interview data, since these activities can be the most problematic in its application to qualitative research. Common objectives in organising and analysing data are:
- To identify key concepts in newly emerging fields
- Structuring an individual's world view from interview data
- To construct a case study
- As an aid to (grounded) theory construction

Although many of the problems of NNV analysis are related more to general issues in the use of interview data, e.g. ethical issues in interviewing, the effort required to learn how to handle data using NNV can cause researchers to forget these more general problems and the need to deal with them. This paper is not, however, concerned with the general problems of interviewing (Alvesson, 2003; David and Sutton, 2004); only those that can affect the final quality of the NNV analysis. Nor is it concerned with the detail of how NNV are used on the PC (an excellent guide to the application of NVivo in qualitative research, which also touches on a number of the issues considered later, is Gibbs, 2002). This paper is about the practical problems of interpretation that arise in applying NNV successfully in interviewing in a global context (with examples) and ways in which these problems can be mitigated by the adoption of particular practices in qualitative data collection or by using NNV in different ways. Its aim is to suggest ways in which NNV practice can be improved so that it leads to a reasonably transparent and consistent analysis, which is seen to conform to, accepted standards of management research. This is particularly timely
given the recent ESRC research project, “Benchmarking Good Practice in Qualitative Management Research,” (Cassell, Buehring, Symon, Johnson and Bishop, 2005) in which good analytical practice was identified as “ranging from demonstrating a systematic, highly procedural approach to emphasising narrative and reflexivity.” (2005, p. 77). Our paper aims to show how good practice in NNV can be developed so that these are achieved. The paper is based on experience of applying NNV in studies of:

- Strategy of Manufacturing Firms
- Nursing Management
- Evolution of e-commerce
- Logistics Industry
- Knowledge management in legal firms
- The place of IT in Design Curricula
- International legal firms

Examples will be drawn from these as appropriate. As already noted, Gibbs (2002) discusses issues that are related to some of those that we consider, e.g. the “tidying up” of grammatical solecisms. On the whole, though, little appears to have been written about the problems with which we are concerned.

1.2 Layout of the paper

The remainder of the paper covers:

- Criticisms of NNV
- Types of users
- Potential impact of interviewees’ motivations on data quality

Specific problems that occur in the application of NNV: a particular focus will be on the fact that interviewing is increasingly being undertaken in a multinational or even global context. Ways of dealing with these issues are then considered and conclusions drawn about good practice in the use of NNV and ways in which these might be encouraged through extensions to the software.

2. Criticisms of NNV

NNV have been the focus of much criticism (cf. Crowley et al, 2002; Webb, 1999). One criticism is that they can cause the loss of the richness of qualitative data (Silverman, 1993; Gilbert, 2002; Seidel and Kelle, 1995). At the opposite extreme, others complain that the effort of coming to grips with them drives researchers to pursue their qualitative research in stereotyped and unilluminating ways, or worse still to use a mechanistic approach to the analysis and presentation of their data (David and Sutton, 2004; Morison, 1998). This danger is perhaps most acute for new researchers such as doctoral students who are struggling to come to grips with the mechanics of applying the programs.

3. Types of users

In the remainder of the paper we will consider three, somewhat stylised, groups of users who can be expected to use NNV in somewhat different ways and to encounter rather different problems in doing so. The groups are shown in Table 1. The three groups vary widely in resources available. However, in line with the existing situation in business research, we assume that none of the parties has more experience of NNV use than a relatively short course concentrating mainly on the mechanics of NNV use and, perhaps, their use in the analysis of mail surveys.

<table>
<thead>
<tr>
<th>Type of Application</th>
<th>Main Researchers</th>
<th>Support</th>
<th>Resources Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major funded research project</td>
<td>Principal Investigator(s) (experienced researchers, not experienced in NNV use)</td>
<td>RA(s) (inexperienced in research, inexperienced NNV users)</td>
<td>RA time, money, NNV training courses, mailing lists/discussion groups</td>
</tr>
<tr>
<td>Specialist research project</td>
<td>Individual internationally recognised researcher</td>
<td>Possibly an RA or attached doctoral students (inexperienced NNV users)</td>
<td>Limited time and money, NNV training courses, mailing lists/discussion groups</td>
</tr>
<tr>
<td>Doctoral research</td>
<td>Research student (inexperienced in research, inexperienced NNV user)</td>
<td>Supervisors (experienced researchers, not experienced in NNV use)</td>
<td>Research student + supervisor guidance, limited finance, NNV training courses, mailing lists/discussion groups</td>
</tr>
</tbody>
</table>

Major research projects are typically either funded through a research grant or as contract research. In either case there will be a “sponsor”. Although with general research council grants the requirements of the “sponsor” may be little different from those of the academic business research community, there are likely to be rather more requirements placed on the researchers if the project has been funded under a research...
council initiative, e.g. to refer in published reports to the initiative’s aims. In the major funded research project the RAs probably conduct most of the interviews and will have the major responsibility for the detail of NNV application. In the second type of project, we envisage someone with an international reputation in a relatively specialist area, e.g. Investment in Taiwan. Such specialists will sometimes have research assistants; they are more likely to have associated doctoral students. For doctoral researchers the principal resources as always, are themselves. Mastery of NNV analysis will be one of the objectives of their doctoral training.

4. Respondent motivation

An important issue in interviewing is the motivation of respondents. It is by no means a universally acknowledged truth that managers and other senior figures in the world of organisations have a strong and disinterested commitment to the furtherance of business research. Accordingly, it is fruitful to consider why individuals participate in interviews (Alvesson, 2003, pp. 15-17). The different motivations for participating lead to different sets of associated problems with the neutrality, consistency, completeness and quality of interview data. There is a temptation, however, once the data have been transferred into NNV to cease to refer back to the recordings/transcripts of the interviews: NNV can, then, cause researchers to forget questionable aspects of their interview data such as the attitude of the interviewee. Of course, where RAs or doctoral students have undertaken the interviewers the PIs may have little idea of the motivation and attitudes of interviewees.

4.1 Reasons for being interviewed and issues associated with them

Common reasons for agreeing to participate in a research interview are:

4.1.1 Edict from on high

Higher-level managers may have granted access to the interviewee. Often the interviewee has not been consulted and has had no explanation of the reasons for senior management’s support of the research. Inclusion in the process can appear to the interviewee to have potential consequences for the existence/form of their job. This may lead to lack of cooperation, obfuscation or disinformation or parroting of “the party line”, i.e. what it is believed that senior management wishes them to say.

4.1.2 Presentation of self

Many, perhaps most, managerial staff need to make project presentations. In a world of increasing employment mobility preparation for job interviews is important. Consultants, internal or external, are similarly continually involved in such presentations. An important component of most of these presentations is the ability to present oneself and one’s experience positively and convincingly at a job interview. Participation in a research interview offers the chance to construct or revise the narrative of one’s personal experience. Obvious dangers of such interviews are: the rewriting of the history of projects usually to emphasise the importance of the respondent’s role or their managerial competence or the scapegoating of others for project failures (cf. Butler et al., 1991)

4.1.3 Presentation of organisation

Interviewees, especially at senior level, may wish to publicise through the research interview strategic initiatives by their organisation or discreetly lobby (government or stakeholder groups). Dangers in such interviews include sanitising of the organisation’s behaviour for the public domain, e.g. pricing initiatives aimed at reducing competition, misrepresentation for competitive advantage, stockmarket impact.

4.1.4 Making sense of one’s own experience

Respondents may simply wish to reflect on their own experience with the aid of an outside perspective offered by the researcher, e.g. in connection with some management development programme. Problems with such interviews include retrospective rationalisation and simplification of that experience.

5. Identifying problems in the application of NNV

Table 2: Problems of NNV use in management research

<table>
<thead>
<tr>
<th>Stage</th>
<th>Problems associated with stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewing</td>
<td>Recording difficulties</td>
</tr>
<tr>
<td></td>
<td>Organisational factors</td>
</tr>
<tr>
<td></td>
<td>Temporal factors</td>
</tr>
<tr>
<td></td>
<td>Individual differences in the ability to recall past events and to</td>
</tr>
<tr>
<td></td>
<td>communicate them to the researcher</td>
</tr>
<tr>
<td></td>
<td>Cultural issues about what may be discussed</td>
</tr>
<tr>
<td>Transcription</td>
<td>Jargon</td>
</tr>
<tr>
<td></td>
<td>Correct transcription</td>
</tr>
<tr>
<td></td>
<td>Understanding spoken English, especially when it is not the</td>
</tr>
<tr>
<td></td>
<td>speaker’s first language</td>
</tr>
</tbody>
</table>
5.1 Interviewing problems

As noted previously, in major research projects the interviewing is likely to be done mainly by the RAs; in specialist projects it is likely to be split between specialist researcher and any RAs attached to the project; and, in doctoral research the interviews will be carried out by the doctoral student. Various generic factors, aside from those of individuals’ motivation, influence interview data gathering and, in turn, impinge on NNV analysis. The most obvious are the problems of understanding recordings of spoken interviews because of, for example, the disjointed nature of ordinary speech, repetition, unfinished statements and often their poor structure. All these problems are exacerbated by the lack, in the recording, of the non-verbal clues provided in the interview situation. There can be difficulties in recording data in the interview process. The unacceptability of tape recording of interviews in some countries is one instance. The more mundane difficulties of ensuring a good tape recording are well covered in standard research texts. There are, of course, ethical issues associated with the recording of interviews (David and Sutton, 2004). In practice, there can be more difficulty with the recording of telephone interviews in that some of the clues available to the face-to-face interviewer are missing in a telephone interview.

Organisational factors are another significant source of difficulty. The most extreme example is perhaps the “our organisation does not participate in management research”. However, the degree of “policing” of interviews, especially of lower level staff, exerted by the organisation can also pose problems. A frequently encountered difficulty is that access to particular groups of staff, e.g. those in overseas subsidiaries. The sheer cost of setting up interviews in other countries may preclude carrying them out if the organisation is unwilling to facilitate consecutive interviews rather than piecemeal visits: telephone interviews, if feasible, do have a very different “flavour” from those conducted face to face. Temporal factors - the effects of events, or opportunities and threats that are particularly salient at the time of interview—obviously play a role. Interviews are genuinely (rather than as an excuse) truncated or postponed because of unexpected initiatives by competitors, the need to prepare replies to parliamentary questions, unfolding strategic events, the need to present analyses to senior management and so on.

There are individual differences in the ability to recall past events and to communicate them to the researcher and, as already pointed out, these may be exacerbated by the motivation of the respondent (Alvesson, 2003). As Alvesson (2003, p.14) points out, “It is important not to simplify and idealise the interview situation, assuming that the interviewee . . . primarily is a competent and moral truth teller, acting in the service of science and producing the data needed to reveal . . . the ‘facts’ of the organization.” In our experience, this poses particular problems in strategy research, which almost always involves a historical perspective. A further complicating factor is cultural issues about what may be discussed. Different cultures have different “taboos” about what may be discussed and how. These may result from differences in professional culture, e.g. the high degree of circumspection that is natural to those operating in the legal field. They may result from different intra-organization cultures and differing perceptions within the organization of the nature of the subject of the interview and the purpose of the interview itself (Miles, 1970). They may arise form differences in national culture or religious culture, e.g. in discussions of Muslim banking. They may be a result of different attitudes to tax authorities or result from stock exchange restrictions on disclosure.

5.1.1 Some solutions to interviewing problems

Table 3 gives some suggested solutions to problems of interviewing for each of our three project types. It presupposes that both RAs and doctoral students are familiar with the prescriptions of standard research methodology texts. A few comments on it are in order. We assume that PIs and specialist researchers have (inter)nationally recognised levels of expertise in the field; RAs and doctoral students do not. PIs and PhD supervisors need to ensure that doctoral students develop appropriate understanding of the relationships within the field of research and of its

<table>
<thead>
<tr>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compensating for interview deficiencies.</td>
</tr>
<tr>
<td>Structuring: defining codes and assigning codes</td>
</tr>
<tr>
<td>Different cultural norms/institutions</td>
</tr>
<tr>
<td>Detecting and compensating for distortions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporting problems</td>
</tr>
</tbody>
</table>
jargon. Access problems plague management research. It is often only possible to resolve these through the intervention of a PI or PhD supervisor. The need for such interventions can be reduced by the provision of training in setting up interviews, which almost always requires arrangement by telephone. Such training is conveniently provided as an adjunct to a course on telephone interviewing, which is likely to be needed for securing supplementary material if for no other reason.

Table 3: Some solutions to interview problems by project type

<table>
<thead>
<tr>
<th>Type of Project</th>
<th>Participant</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major Project</td>
<td>RAs</td>
<td>Immersion in the field, Testing of recording equipment, Training in non-electronic recording methods, e.g. mind maps, Careful listening to interview tapes, Liaison with PIs over major points emerging</td>
</tr>
<tr>
<td></td>
<td>PIs</td>
<td>Monitor access problems and mitigate them, Listen to recordings of interviews, Reorganise interviews as necessary</td>
</tr>
<tr>
<td>Specialist Research Project</td>
<td>RAs (if any)</td>
<td>Close liaison with lead researcher</td>
</tr>
<tr>
<td></td>
<td>Lead researcher</td>
<td>Close liaison with RAs, Monitoring of relationship between PhD students’ work and overall project</td>
</tr>
<tr>
<td>Doctoral Research</td>
<td>Research student</td>
<td>Immersion in the field, Training in non-electronic recording methods, e.g. mind maps, Close liaison with supervisors over the setting up of and progress of interviews, Careful noting of parts of interview not understood.</td>
</tr>
<tr>
<td></td>
<td>Supervisors</td>
<td>Facilitation of access, Monitoring of transcription, Checking of student’s developing understanding of the field</td>
</tr>
</tbody>
</table>

Not all interviewees will agree to the interview being recorded. For this reason, RAs and PhD students need training in non-electronic means of note taking, e.g. “Mind Maps” (Buzan, 2004), that can be practised in the interview. Some of these methods could be used potentially to record non-verbal cues arising during the interview. In addition to the project specific recommendations of Table 3, there are some more general suggestions that can be made. In major or specialist research projects, we have found it convenient to record interviews using conventional cassette tape recorders. These can give at least 90 minutes of recording, the cassettes are easily copied using conventional equipment and all involved in the project can play back the interviews whilst stuck in traffic jams! Unfortunately, such equipment is becoming harder to find. A further “wrinkle” is to play back the tape for transcription on the machine on which it was originally recorded. Playing it back on another machine will certainly not improve the quality; on some occasions it may be markedly worse.

5.2 Transcription problems

In the major funded research project the RAs will usually carry out the task of transcription of interviews. As noted by Fielding and Lee (1998) this is a key activity. Loxley (2001) discusses the importance of close liaison (which is sometimes difficult to achieve) between the PIs and the RAs at the early stage of formulation of the analysis. Equally, there can be problems in understanding what respondents mean: jargon problems (the researcher may well not know what particular terms mean); language problems (in a global context, either interviewer or interviewee may be operating in their second language and the controls available through back translation in survey research are not possible in interviews).

There is an increasing tendency in multinational research to employ RAs or use PhD students who can conduct interviews in languages other than English, which implies a need for translation of the material gathered. However, a need for “translation” also arises when the interview has been conducted in English when the interviewee is not completely fluent in that language. Consider, for example, the following quote from an interview:

“We believe that we have a big impact on our industry itself. We are part in many organisations and give feedback of our ideas into this”.

There would seem to be two obvious ways of dealing with such problems:

1. Code the text according to the researcher’s interpretation, e.g. Contribution to other
companies’ capabilities through working with them; Impact of the company on the industry;
2. (Freely) translate the text and use that translation in the NNV analysis (with the original being attached as a memo).

Whichever approach is adopted, we believe that it should be explicitly recognised in the final research report that significant, and not necessarily, uncontroversial adaptation of the text has taken place. This, may, however, raise issues of confidentiality. Similar problems arise in connection with transcripts generated from with native English speakers. At the simplest level, these may involve ‘er’s missing verbs, incoherent tenses, and so on. We would endorse Gibb’s (2002) view that these should be simply tidied up for the purpose of analysis with the original attached as a memo. We would advocate that the same approach should also be applied to pronouns, which may be ambiguous to conceal prejudice (cf. Fielding and Lee, 1998), or simply ambiguous because of inevitable ellipsis in a lively interview. Because of the potential impact of motivation we would also suggest that, at least where the motivation appears unusual or potentially problematic, that a suitable memo should be attached to the transcript.

A further problem, already raised above in the discussion of interviewing problems are the non-verbal clues, not only are they difficult to record, but even if recorded difficult to transcribe. As Gibson et al, (2005: section 1.3) highlight in their discussion of uses of audio/audiovisual data in qualitative research:

“The written representation of speech through transcription is a poor substitute for the data to which it pertains; even complex representation structures . . . are not adequate for picking up all the relevant inflections . . . The result is that it can be extremely hard to communicate the complexity of meaning embedded within a given interactional event through transcription.”

Implicit in the above suggestions is that they need to be implemented by the interviewer as soon as possible after the interview. In our experience, there are usually gaps in interview programmes that will allow RAs or PhD students to do this. Ideally transcription ought to be done by the RAs, though in major research projects at least a specialist transcriber may do the first transcription. In such cases the aim should be to get the “first cut” transcriptions back to interviewer as soon as possible. There are, however, problems that arise in transcription that are not easily dealt with by relatively inexperienced RAs or PhD students. There can be jargon problems; the researcher may well not know what particular terms mean and indeed they may mean different things in different contexts, e.g. Internet Marketing or Information Management for IM. Homophones may be difficult to interpret, e.g. the UK information systems quality system ‘TickIt’ being transcribed as ‘ticket’, ‘ASEAN’ being mis-transcribed as ‘Asian’. PIs and single researchers can be expected to know such terms, RAs not. Our suggestions for avoiding transcription problems for the 3 types of project are given in Table 4 below.

<table>
<thead>
<tr>
<th>Table 4: Some solutions to transcription problems by project type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Project</strong></td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Major Project</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>RAs</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>PIs</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Specialist Research Project</td>
</tr>
<tr>
<td>Lead researcher</td>
</tr>
<tr>
<td>Doctoral Research</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Supervisors</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
5.3 Problems of NNV analysis of interview data

5.3.1 Compensating for interview deficiencies

The problems of gathering data through interview, listed above, can translate into problems of NNV analysis. The interviews may have produced “one respondent” views. Imperfect recall can have resulted in conflicting accounts of past events by respondents. The interviews may have reproduced biased or inadequate worldviews of respondents combine these. These, of course, are not problems that can be remedied at the analysis stage; some re-interviewing (to amplify respondents’ views) or additional interviewing (to provide another’s perspective) can be essential.

5.3.2 Structuring: defining codes and assigning codes

The issue of how to define codes and the extent to which they should emerge from the data is a longstanding question in NNV analysis (cf. Gibbs 2002). Our contention is, however, that efficient NNV analysis and the growing professionalisation of management research, as evidenced by the AIM initiative, require (pace Glaser and Strauss, 1967) that the researcher bring to bear some preliminary structuring of the research area at the start of the NNV analysis. It is extremely rare that no worthwhile concepts can be derived from the literature to address what almost by definition will be a relatively new subject. Indeed the requirements for literature reviews in research grant proposals and PhD theses and ever more extensive sections dealing with the literature in leading management journal papers suggest that this is taken for granted. In our experience, the assumption by research students that their NNV coding has to be derived completely from the data because the field is “so new” leads to poorly structured NNV analyses.

Of course, the fact that the analysis commences with a node tree derived from the literature does not mean that that tree is sacrosanct. It must be regularly reviewed. A further aspect of the structuring of the data is the assignment of codes. Two principles should, we suggest guide practice here: 1) to submit coding to crosschecking by others; 2) coding should be continuously checked (Loxley 2001; Silverman 1993).

5.3.3 Different cultural norms/institutions

A factor that adds considerably to the problems of coding, especially in international research are different cultural norms/different institutional roles in different countries. An example is the Code Napoléon which underpins very different legal forms and practices from those of the anglo saxon common law tradition in the many countries whose legal system is based upon it. The solution again has to be thorough immersion in the context within which the research is embedded.

5.3.4 Detecting and compensating for distortions

We noted earlier that the differing motivations of respondents can to suppression of certain issues or a sanitised account of them. This deliberate distortion is different, in principle, from the inadvertent distortion that occurs when a respondent simply forgets or cannot realistically provide information on something that happened in the past. Signs of the latter problem include: the respondent is only willing to talk about their own function (this can be used often as a lever to enable the researcher to talk to other functions); the respondent has no hands on experience of the subject; the respondent is recounting projects for which they took over responsibility late on. Interviewing several people about, say, different aspects of a decision can help to reduce this problem. A similar effect can be achieved in some types of research by interviewing both client and consultant. Both types of distortion obviously affect the integrity of the NNV data. It is, therefore, good practice to check individuals’ accounts against sources such as databases of company information, the quality press or the relevant professional trade journals. The two latter sources are also very useful in dealing with deliberate distortion but, in this case, they need to be supplemented by “gossip” within the industry concerned. Because this type of investigation is more time consuming, it is useful to have produced memos indicating that it is needed at the transcription stage. Our suggestions as to how some of the problems of analysis can be mitigated are given in Table 5.
Table 5: Some solutions to analysis problems by project type

<table>
<thead>
<tr>
<th>Type of Project</th>
<th>Participant</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major Project</td>
<td>RAs</td>
<td>Immersion in the field</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identify where re-interviews/extra interviews needed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Supplementation of PI’s initial structuring based on literature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agree as part of team (PIs dominant) preliminary structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Report by email problems during pilot to PIs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Report to full team /suggest revisions after first analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAs compare codings between themselves</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Document revisions to the node tree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Check transcripts against “alternative evidence” where possible paying special attention to interviews that may have involved deliberate distortion</td>
</tr>
<tr>
<td></td>
<td>PIs</td>
<td>Preliminary structuring of research area to generate initial node tree. Team (PIs dominant) agrees preliminary structure.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sample checks on RA codings.</td>
</tr>
<tr>
<td>Specialist Research Project</td>
<td>RAs (if any)</td>
<td>Immersion in the field</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Review transcripts of interviews to ascertain where re-interviews/extra interviews needed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agree preliminary structure and revisions with lead researcher.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agree codings with lead researcher.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agree with lead researcher “alternative evidence” to be gathered and gather it.</td>
</tr>
<tr>
<td></td>
<td>Lead researcher</td>
<td>Preliminary structure of node tree.</td>
</tr>
<tr>
<td>Doctoral Research</td>
<td>Research student</td>
<td>Immersion in the field</td>
</tr>
<tr>
<td>Supervisors</td>
<td>Checking of student’s developing understanding of the field</td>
<td>Review gaps in interview findings</td>
</tr>
</tbody>
</table>

5.4 Reporting problems

NNV analysis can produce in the final research report extraordinarily boring quotes, devoid of insight. This problem is exacerbated where quotations from interviews is used as a substitute for analysis (Gibbs, 2002). There is a real danger, especially in doctoral theses, that the research report becomes large slabs of quotation interspersed with relatively unilluminating comment. Although if our recommendations on translation are followed such texts should at least be presented in reasonable English, this is far from guaranteeing that they will be found particularly interesting to the reader. The interviewer, of course, may see them differently because they can judge the importance of the comment in its context. A quote such as: ‘We need to understand our customers better and find out what it is they find of value and what it is they don’t and what other kinds of things we can do to make their life better and easier’ may in fact capture a moment where the interview was able to launch into an illuminating discussion of marketing strategy; to someone unconnected with the research it merely appears banal. If any worthwhile meaning can be extracted from it, it will only be because the importance of the quote to understanding the phenomenon under discussion has been explained.

The above discussion really relates to the “illuminating quotes” that are usually much more apparent to the researcher than the dispassionate observer! There is, however, another, in principle, deeper reason for providing them: they document the way the researchers have analysed their qualitative data. Of course, the problem here is that only a small, and probably unrepresentative, fraction of the material can be cited. In short, our view is that quotations should be used only sparingly and appositely. It is usually better to provide a synoptic view that brings out the commonalities between, say, different cases and to note any significant differences between cases.
Table 6 gives our suggestions as to how the quality of the reporting of NNV analysis can be prepared. In this case there are a number of recommendations that apply to any type of research project.

Table 6: Some solutions to analysis reporting by project type

<table>
<thead>
<tr>
<th>Type of Project</th>
<th>Participant</th>
<th>Action</th>
</tr>
</thead>
</table>
| All projects                     | All involved in writing the research report | Base the report on a “frozen” version of the NNV database  
Formulate a clear “house style” (section headings, etc) for the report  
Devise, and revise as necessary, a clear outline structure for the research report that relates to the research questions to be addressed  
Continually ask which parts of the NNV analysis relate to this aspect of the outline?  
Continually ask the questions is this the logical place for this passage?  
Have I provided the target reader by this point with enough information to understand what I have written?  
Try drafts sections out on others involved in the project  
Try later drafts out on researchers not involved in the project.  
Unless the topic is extraordinarily arcane, try out report sections on friends who know nothing of the subject  
Carefully check the final report for spelling and grammatical errors and adherence to the “house style” |
| Major Project                    | RAs           | Agree as part of team (PIs dominant) the structure of the report and who takes responsibility for which sections  
Check and correct penultimate version |
| PIs                              |               | Define the structure of the report and who takes responsibility for which sections  
Check individual sections as written and penultimate version of entire report |
| Specialist Research Project      | RAs (if any)  | Agree with lead researcher the structure of the report and who takes responsibility for which sections |
| Lead researcher                  |               | Agree with RAs the structure of the report and who takes responsibility for which sections |
| Doctoral Research                | Research student | Devising a structure for the thesis  
Preparing the thesis |
|                                  | Supervisors   | Commenting on thesis structure  
Read individual chapters (and revisions) and entire thesis  
Checking student has displayed in the thesis understanding of the wider field within which the research is located  
Critical review of the structure and argument of the thesis |

6. Conclusions

A number of considerations arise from the arguments above. We organise our conclusions under the headings:

- Generic good practice in applying NNV
- Some desirable extensions to the NNV software to facilitate better analysis
- Interpretivism and NNV research
- Specialist research projects
- Supervisors’ duties.

While we have focused throughout on the application of NNV in particular, much of our discussion is equally relevant for the application of other CAQDAS tools, e.g., Atlas.ti. However, users of other tools should be aware that the issues we raise and the solutions we suggest, where they relate to the technical side of NNV, are specific to that tool and may be of different relevance or require adaptation when applied to other tools.

6.1 Generic good practice in applying NNV

Gibbs’ (2002, p. 13) points to the danger that NNV are used to impart a ‘gloss of rigour’ to research. We would contend that this danger is receding as more management researchers become familiar with the pitfalls that can occur in the application of NNV and with how these can be avoided. Since NNV, however, beguilingly apparently offer the chance to analyse many more data far more rapidly, there is a real danger that their use can lead to significantly poorer quality analysis of what, ostensibly, are more firmly grounded
research findings. Many of the problems of applying NNV successfully are, of course, more general problems of qualitative business research but the practical difficulties of coming to grips with the technical side of NNV and translating the interview data into NNV often exacerbates and conceals the need to consider, and mitigate these problems. We presented above in some detail approaches that can be employed to improve the effectiveness of NNV use. These can be summarised in the more general precepts below.

6.1.1 Understand the research field thoroughly

Especially in international research, researchers need to have a thorough understanding of their field. This has clear implications for the formation of research assistants and PhD students.

6.1.2 Pay attention to the actual data at all stages

It is crucial to listen to the interviews, read the transcripts, etc. This has clear implications for the size of research project that individual or small groups of researchers can deal with and for division of labour within major research projects.

6.1.3 Face up to the fact that interviews are rarely complete

Newer researchers, in particular, therefore need training to set up going back to respondent if necessary. In larger projects, it is helpful to work on a snowball-sampling basis; with further interviews being set up if preliminary analysis suggests this is warranted. For doctoral students with limited time it is worth enquiring whether there is anyone else who can be interviewed in the same visit (often a measure of successful interviews).

6.1.4 Start with a structure and continually revise it

A good initial node tree based on the literature will lead to a more rapid and deeper NNV analysis provided that it is continuously revised and reviewed. In larger projects, this should be done by formal sessions of the entire team as well as adaptation in periods between such sessions.

6.1.5 Continually check the coding

As suggested by Loxley (2001) and Silverman (1993) coding should be kept under continuous review. It is essential that others review coding. This has implications for resource allocation in major research projects (where often different aspects of the research will be allocated to different RAs for first cut coding) and for doctoral researchers who can perhaps negotiate peer review of coding on a quid pro quo basis. Both PIs and supervisors need carefully to monitor the quality of coding.

6.2 Some desirable extensions to the NNV software to facilitate better analysis

The importance of non-verbal cues during the interview was discussed earlier. It would be useful if NNV provided a way of recording these non-verbal cues explicitly rather than doing so through memos. Gibson et al. (2005) suggest for example, the embedding of digital data files. The emphasis placed above on continual revision of NNV model structure and codings implies that the NNV model undergoes continuous change. There is a need to keep track of such changes, at least of a more major nature such as the transfer of node subtrees from one node to another (in case they turn out not to be a good idea). This implies that rather than update the existing model thus losing the details of its past form that a new, revised model should be created from it. Eventually, the final report will need to be based on an agreed NNV model, though additions may still be made to another model with an eye to future research. It is difficult for the individual student researcher to observe this requirement. In team research it is clearly more difficult; the team need to be clear which is the most up-to-date version of the NNV model. This would be helped if NNV supported more directly versioning in the way that is embodied in software for supporting software development projects.

We argued that the amount of quotation from NNV documents even after “translation” ought to be severely limited. One way of doing this would be to allow hyperlinking to relevant documents in NNV databases so that readers interested in the detail of a specific analysis could pursue it in depth. Of course, considerations of confidentiality may limit the universal applicability of this idea. It could, though, be partially achieved within the electronic version of the thesis by making use of the hyperlinking facility of Microsoft Office.

6.3 Interpretivism and NNV research

NNV would probably be viewed by most management researchers as essentially based in a positivist philosophy. However we would argue that good NNV analysis is more interpretivist than might be thought. There can be a need to interpret much of the raw data before it is manipulated within NNV. This is more likely when research is being conducted on a transnational basis since additional complications arise through differences
in language and culture. What we have dubbed “translation” essentially involves trying to understand what interviewees meant. Likewise, we have pointed to the importance of understanding respondent motivations in interpreting their statements and judging the need for supplementary research. As researchers become more familiar with NNV it is to hoped that they will be seen as potentially beneficial in interpretivist research also.

6.4 Specialist research projects

A notable feature in the discussions above is the relatively good performance of specialist research projects where an RA is attached. This is because of the close collaboration of RA and lead researcher that is possible (Loxley, 2001). This, of course, is only feasible if the lead researcher has the time to devote to the research. It would seem worth examining the potential for strengthening this type of management research by the provision of relatively modest funds for a research assistant and for teaching relief for the lead researcher.

6.5 Supervisors’ duties

It will not have escaped the notice of PhD supervisors that in our discussion of their role it is considerably expanded beyond what may have been the case in the past. This, in part, reflects the growing importance being attached to quality assurance in doctoral studies. More importantly, however, effective qualitative analysis using NNV requires that supervisors pay even greater attention than in the past to the extent to which their students are developing understanding of their field. The downside of NNV analysis for supervisors is that they must monitor carefully how students develop and analyse their NNV model. And, of course, they need to encourage students not only to develop a good thesis structure and sound argumentation but also to try to ensure that the student reports the NNV analysis in a way that enhances, rather than detracts from, the thesis.

References


Applying Multidimensional Item Response Theory Analysis to a Measure of Meta-Perspective Performance

K. Michele Kacmar¹, William L. Farmer², Suzanne Zivnuska³, L. A. Witt⁴
¹ Department of Management and Marketing, University of Alabama, USA
² FedEx Express, Memphis, USA
³ College of Business, California State University, USA
⁴ University of Houston, USA

Abstract: The authors introduce a scale to measure meta-perspectives, my view of your view of me, about one's performance in an organizational setting. Applied to the performance appraisal process, this perspective allows the authors to investigate how employees think their supervisors view their performance. Meta-perspectives thereby enrich our understanding of the relationship effects inherent in the performance appraisal process. Due to the desirable properties of item response theory (non-sample specific item parameter estimates), a multidimensional item response theory (MIRT) model was applied to the data. This allowed for the simultaneous estimation of dimensionality and item threshold values. Data collected from 1,255 full-time workers in two different organizations reveal that the items did not lie along a unidimensional continuum, but that three dimensions underlie the proposed scale: employee perceptions of the supervisor's view of employee work ethic, work product, and self-regulation. The authors offer suggestions for refinement of the scale and future research.

Keywords: Item response theory, scale development

1. Introduction

If called upon by another to describe oneself, few individuals would have difficulty responding. Most individuals have a strong sense of self, frequently referred to as one's self-identity. Self-identity is the view of oneself created and maintained over time through inter- and intra-personal experiences that define what the self is like (Schlenker, 1986). Self-identity is formed by showing oneself to be a particular type of person both publicly, through self-presentation and self-disclosure, and privately, through introspection and contemplation of oneself (Schlenker 1984, 1985).

One's self-identity is how one sees oneself (i.e., my view of myself). The identity that results from acting on the experience of another's actions, as illustrated by the dyad, is a “meta-perspective” (Kenny and Albright, 1987; Patterson, Churchill, Farag, and Borden, 1992) – or how I think you see me (i.e., my view of your view of me). Given that an individual defines his or her own self-identity, one's view of oneself is generally not open to refutation. However, the same cannot always be said for one's meta-perspective, as an individual can never be truly sure he or she knows why others act as they do. Instead, an individual understands another's actions based on his or her own perceptions. Frequently, the understanding that an individual has of another's behaviour is incorrect, as the individual simply projects his or her own reasons for acting a certain way onto the other. Unaware of the perceptual error, the individual continues to react to the other's behaviours based on false assumptions of why the behaviours were enacted. These false assumptions can lead to mismatched perceptions that can lead to a spiral of mismatched perceptions, resulting in a less than effective communication experience. An example may make this more concrete.

Pat reports to Chris. Pat sees himself as possessing strong leadership potential and aspires to be promoted to management. Thus, at
every opportunity Pat engages in leader like behaviour such as setting deadlines for colleague’s projects and reorganizing the office to make it more efficient. Pat fully admits that enacting these behaviours some times gets in the way of finishing his assigned duties, but believes his actions are appropriate given the continuous feedback he receives from Chris about his actions. Chris recognizes Pat’s "need to lead," but attributes it to being lazy rather than a strong leader. Chris sees Pat’s behaviour as a blatant attempt to get out of doing his assigned work. Chris, regularly reminds Pat that his extra-role behaviours - even when they are on target - take away from his in-role behaviour upon which he will be appraised. Appraisal time is upon them. Pat is confident that Chris will provide him a strong evaluation and place his name in the promotion pool so that he can apply his management skills full time. Unfortunately, Chris rated Pat low because he continually misses deadlines by engaging in distracting “management-like behaviours” rather than finishing his assigned work. Pat thought Chris viewed him as a strong leader when in reality Chris saw Pat as a slacker. While both thought they were clearly communicating their positions to the other, a meta level miscommunication occurred.

The focus of the present study is on the meta-perspective described above. The decision to study meta-perspectives was based on our belief that many of the miscommunications that occur in the workplace can be explained by inaccurate perceptions about others’ behaviours. Supervisors and subordinates modify their own behaviour based on limited and frequently inaccurate information about why the other is acting the way he or she is. As each person continues to behave based on a flawed understanding of who he or she is to the other person, rather than on the other’s true beliefs, both individuals may get caught in an escalating series of miscommunications. Thus, gaining an understanding of how meta-perspectives are formed and how they contribute to ineffective dyadic relationships in the workplace may help us find a way to create more positive interactions.

This research represents a first step towards better understanding the occurrence of meta-perspectives. Our initial task was to develop a scale to measure meta-perspectives in an organizational setting. Although there are many different organizational issues in which meta-perspective can be applied, we elected to focus on performance ratings, as this is one of the most critical human resource issues organizations face. Organizations cannot be competitive without employees who can effectively perform their jobs. Accordingly, the primary purpose of the present study was to assess the dimensionality of our newly developed meta-perspective performance scale via multidimensional item response theory analysis (e.g., Ackerman, 1987) prior to assessing its construct validity and testing hypotheses.

2. A social relations explanation of meta-perspective

Scholars have theorized that one’s identity influences how a person behaves and therefore how he or she is perceived by external others (Albert and Whetten, 1985; Ashforth and Mael, 1989). In essence, we all form our personal identity in an effort to answer the question, “Who am I?” (Ashforth and Mael, 1989). To reduce the possibility of cognitive dissonance, we are then apt to implement behaviours that enable us to confirm our self-defined identities.

In the context of a subordinate’s role, if the answer to “Who am I?” is, “I am a conscientious employee,” that identity may well lead to different behavioural patterns than if the answer is, “I am a team player.” For example, the self-identified conscientious employee is likely to focus his/her behaviours on meeting deadlines and producing high quality work. On the other hand, the self-identified team player may be likely to focus on shared leadership and trust building.

The social relations model (Kenny, 1981; Kenny and La Voie, 1984) takes this identity-behaviour relationship a step further, and posits that individual behaviour is contingent upon three relationship factors: (1) an actor effect (consistency of actor’s behaviour across partners), (2) a relationship effect (the unique ways an actor behaves with a given partner), and (3) a partner effect (consistency in the behaviour of the actor’s partner). Thus, if I see person X consistently act in a cheerful, helpful manner towards many people (actor effect), if person X is consistently helpful in interactions with me (relationship effect), and if I consistently ask person X for favours in anticipation of his or her helpful response (partner effect), person X’s tendency is not only to think of himself or herself as helpful, but is also to think that I think he or she is helpful (meta-perspective). Thus, these effects suggest that individuals alter their behaviour based on the subjective perceptions they hold about the perceptions held by their interaction partners; people alter their behaviour based on how they think others perceive them (DePaulo, Kenny, Hoover, Webb, and Oliver, 1987).

Because person perception is argued to be contingent in almost equal proportions upon the
perceiver as well as the person being perceived (Bourne, 1977; Kenny and La Voie, 1984), reaching agreement between the perceiver and the perceived individual is an important determinant of future behaviour. In other words, if a subordinate believes that he or she is perceived as a helpful team member, when in fact the supervisor views the subordinate as overly controlling and bossy, the subordinate may be apt to blithely continue the controlling, bossy behaviors, which the supervisor finds frustrating. Eventually, this mismatch between the two different interpretations of the same behaviour may result in tension between the employee and the manager. If the subordinate is effective at reading the reactions of his or her manager, he or she will realize that a less pushy, more low-key approach to team membership is desirable and will change his or her behaviours accordingly. However, if a mismatch exists but the subordinate does not recognize it, the employee will not change his or her behaviours and an ongoing performance problem may result.

In fact, mismatched perceptions between subordinates and supervisors may be more common than congruent perceptions (Harris and Schaubroek, 1988), and subordinates often expect higher performance ratings than they get in formal reviews (Pearce and Porter, 1986), indicating that their meta-perspectives are incorrect. Thus, many of the problems associated with the performance review process (e.g., defensiveness, conflict, political posturing), could be ameliorated by increasing the accuracy of subordinates’ meta-perspectives. Therefore, our initial efforts in this line of research were directed towards developing an empirical measure of performance meta-perspectives in an organizational setting.

Due to the desirable psychometric properties inherent in multidimensional item response theory (MIRT) (e.g., non-sample specific item parameter estimates), we selected this analysis approach to model our data. MIRT combines features of both unidimensional item response theory (UIRT) and factor analysis (FA), to provide both dimensional and item threshold information. A secondary purpose of our study was to expand the exposure of MIRT methods in an applied setting.

3. Method

3.1 Sample

A survey was administered to two different samples located in the southeastern United States. One organization was a large division of a state government agency, and the other was an electrical cooperative. The employees of the state agency provided a total of 786 usable responses, and the employees of the electric cooperative returned 469 usable responses. Both surveys were administered as part of a larger employee opinion survey conducted at the request of each organization. Given this format, we were not provided complete freedom in construction of the surveys. One restriction placed upon us was that no demographic data be collected, because top management from each organization believed that this information could be used to determine the identity of specific respondents, thus eliminating the anonymity promised. Although specific demographic data are not available, there was considerable diversity among the respondents with respect to gender, race, and age, because 67% of the state agency and virtually every member of the electric cooperative (95%) responded to the survey. With such high response rates, the demographic characteristics of the actual population are reflective of the sample. For the state agency, 14% of the employees were African American, 37% were female, and the average age was 42.1 years of age. The respondents from the electrical cooperative included 5% African Americans, 22% females, and had an average age of 44.6 years of age. When conducting the multidimensional item response theory analyses, we combined the two samples in order to create one dataset with 1255 responses. This was done with the intent of maximizing the stability of derived parameter estimates (Ackerman, 1994), by ensuring a sample more representative of the general population.

3.2 Procedure

State agency. The surveys were delivered to the director of the state government agency. He wrote an endorsement letter that was distributed with the survey via interoffice mail to each employee in the agency. Included with the survey was a return envelope addressed to the researchers. The cover letter directed the respondents to complete the survey on company time and then seal it in the envelope and mail it directly to the researchers. Of the 1175 surveys sent, 786 (67%) were returned within the three-week time period allocated.

Electrical cooperative. The surveys were mailed to the Human Resource Director for the cooperative sample. His assistant distributed the surveys to groups of 15-25 employees assembled in a conference room at the general offices over a three-day period. The respondents, collected by the assistant and mailed to the researchers, sealed the surveys in return envelopes. All employees present at work on any one of the three days of survey administration participated.
providing 469 usable responses out of 495 employees for a response rate of 95%.

3.3 Measures

Our first step was to create items to measure meta-perspectives in an organizational setting. To accomplish this goal, we created meta-perspective items about an important issue in organizations -- performance appraisals. Specifically, we modified generic performance appraisal items (Wright, Kacmar, McMahan, and DeLeuw, 1995) to reflect a meta-perspective by reversing the referent focus of the item. For example, the performance appraisal item that asked a supervisor whether an employee always tried to do things better at work became "I think that my supervisor thinks that I always try to do things better at work" (1 = "strongly agree;" 5 = "strongly disagree"). The items are presented in the Appendix.

3.4 Analyses

We applied multidimensional item response theory analysis (MIRT; Ackerman, 1987, 1992, 1999; McDonald, 2000; Oshima and Miller, 1991; Reckase, 1997; Reckase, Ackerman, and Carlson, 1988; Way, Ansley, and Forsyth, 1989) to the eight meta-perspective items presented in the Appendix. As mentioned in our measures section, these items were modified versions of an existing performance scale (Wright et al., 1995) that captures four key components of successful performance: initiative, diligence, competence, and confidence. Our intent was to develop an empirical understanding of both the construct and item-level measurement characteristics of these items. MIRT models a set of item responses utilizing features of both factor analysis and item response theory (Farmer, 2001). A PsychInfo search of the 1992-2004 abstracts of the literature for either "multidimensional item response theory" or "MIRT" appearing in journal articles published in English revealed only twenty studies. MIRT has been successfully applied in studies regarding measurement issues (e.g., Ackerman, 1996; Douglas, Roussos, and Stout, 1996; Ferrando and Lorenzo, 1998; Janssen, and De Boeck, 1999; Luecht, 1996; Luecht, and Miller, 1993; Miller and Hirsch, 1999; Segall, 2001). However, despite its usefulness and appropriateness, we are aware of only one previous study that applied MIRT to issues of relevance to the I/O literature (Farmer, 2002).

Factor analysis is a "dimension discovery" technique (Brown, 1994) that is typically utilized to reduce a set of measured variables to an interpretable set of fewer "unmeasured" or latent variables. These latent variables are generally hypothesized to underlie the measured variables and typically are of more theoretical interest than the individual items themselves (Cattell, 1966). Though confirmatory methods (including structural equation modelling) have tended to dominate the literature in recent years, exploratory factor analytic techniques still serve an invaluable role in establishing dimensionality of a new measuring instrument (as the one used in this study) or in the absence of substantive theory. Exploratory efforts are usually based on the common factor model, in which the response to any variable in a dataset is modelled as a linear weighted combination of a set of latent variables that are “common” to all of the variables. The remaining variance in each item is accounted for by a factor unique to that variable.

In item response theory (IRT), a set of items is modelled along an assumed latent trait, resulting in an index (item difficulty or threshold) of the item’s position on the latent continuum. In Likert-type data, thresholds are produced for each item category (minus 1). In addition, in models that consider item discrimination, an index of the item’s ability to differentiate those at particular points on the continuum is provided. Whereas in factor analysis multidimensionality is explicitly modelled, commonly used IRT models generally operate under the assumption that a set of items modelled together lie somewhere along a single unidimensional continuum. Though a fair amount of research (Hambleton, 1989) has been done to examine the effects of multidimensionality in a dataset, until recently practical efforts were directed at minimizing the effects of dimensions other than the one explicitly assumed to underlie the modelled items. This situation, along with a general requirement of large sample sizes, has hindered a widespread utilization of IRT methods in most organizational research (Tenopyr, 1994).

MIRT methods assume that more than one dimension underlies the responses to a set of items, and models each item in a continuous multidimensional space. Item thresholds are produced, as in unidimensional item response theory (UIRT) models, as are item discrimination indices. Unlike in UIRT models, which provide one discrimination index per item, MIRT analyses result in the same number of discrimination indices, as there are modelled dimensions. In fact, they serve as the MIRT analogue to factor loadings.

That being stated, the focus of this analysis is to simultaneously model the data dimensionality and the item-level measurement characteristics of the eight survey items. The analysis was performed using Mplus 2 (Muthen and Muthen, 2001). Mplus
provides a number of modelling options that allow for exploratory or confirmatory latent variable and latent class analysis, and for any combination of continuous and categorical measured variable formats.

4. Results

Item inter-correlations are presented in Table 1. For the purpose of this study, the meta-perspective items were modelled as ordinal categorical variables, which was necessary from an IRT perspective. Therefore, the coefficients that are presented are polychoric.

**Table 1: Meta-perspective item inter-correlations**

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>---</td>
<td>.80</td>
<td>.78</td>
<td>.71</td>
<td>.59</td>
<td>.61</td>
<td>.54</td>
<td>.57</td>
</tr>
<tr>
<td>2</td>
<td>.80</td>
<td>---</td>
<td>.84</td>
<td>.73</td>
<td>.63</td>
<td>.62</td>
<td>.55</td>
<td>.56</td>
</tr>
<tr>
<td>3</td>
<td>.78</td>
<td>.84</td>
<td>---</td>
<td>.70</td>
<td>.63</td>
<td>.66</td>
<td>.60</td>
<td>.61</td>
</tr>
<tr>
<td>4</td>
<td>.71</td>
<td>.73</td>
<td>.70</td>
<td>---</td>
<td>.70</td>
<td>.66</td>
<td>.57</td>
<td>.58</td>
</tr>
<tr>
<td>5</td>
<td>.59</td>
<td>.63</td>
<td>.60</td>
<td>.70</td>
<td>---</td>
<td>.67</td>
<td>.62</td>
<td>.60</td>
</tr>
<tr>
<td>6</td>
<td>.61</td>
<td>.62</td>
<td>.66</td>
<td>.67</td>
<td>.76</td>
<td>---</td>
<td>.58</td>
<td>.61</td>
</tr>
<tr>
<td>7</td>
<td>.54</td>
<td>.55</td>
<td>.60</td>
<td>.57</td>
<td>.62</td>
<td>.60</td>
<td>---</td>
<td>.65</td>
</tr>
<tr>
<td>8</td>
<td>.57</td>
<td>.56</td>
<td>.61</td>
<td>.58</td>
<td>.60</td>
<td>.61</td>
<td>.65</td>
<td>---</td>
</tr>
</tbody>
</table>

*Note: Polychoric coefficients; N = 1,255.

Based on a scree test plot of eigenvalues (Cattell, 1966), we determined that three underlying dimensions were probably accounting for the variance in the correlation matrix. Due to the exploratory nature of this study, an exploratory model-testing paradigm was used. A number of models (ranging from 1- to 4-dimensional solutions) were tested using a chi-square difference test to arrive at a model that fit the data. In addition, RMSEA and RMSR indices were used in the evaluation of model fit. Model testing results are presented in Table 2.

**Table 2: Results of model testing procedure for 1- to 4-dimensional solutions**

<table>
<thead>
<tr>
<th>Item</th>
<th>1-dim</th>
<th>2-dim</th>
<th>3-dim</th>
<th>4-dim</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>( p )-value</th>
<th>RMSEA</th>
<th>RMSR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>433.1</td>
<td>20</td>
<td>&lt;.001</td>
<td>.128</td>
<td>.156</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>164.5</td>
<td>13</td>
<td>&lt;.001</td>
<td>.096</td>
<td>.058</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26.22</td>
<td>7</td>
<td>.001</td>
<td>.047</td>
<td>.015</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.857</td>
<td>2</td>
<td>.391</td>
<td>.009</td>
<td>.003</td>
</tr>
</tbody>
</table>

Note: N = 1,255. RMSEA = Root mean square error of approximation. RMSR = Root mean square residual.

Using a combination of model fit indices, we determined that the 3- and 4-dimension solutions most closely approximated the underlying structure of the data. After further evaluation, in which the criterion of interpretability was used as the primary criterion, we decided to use a 3-dimensional solution. Further, rotational analysis (both orthogonal and oblique) suggested (mean inter-factor correlation was .693) that allowing the dimensions to correlate was appropriate.

The results of the MIRT analyses are presented in Table 3. The first three columns represent the obtained factor loadings of the items. As illustrated, the first four meta-perspective items load most highly on the first dimension. In succession, items 7 and 8 comprise a second dimension (though item 7 also demonstrates a low loading on the third dimension); and the third dimension is most closely linked to items 5 and 6. The next three columns are the loadings transformed to a logistic metric (Muthen and Muthen, 2001) to allow for an alternate interpretation of the item discrimination values. Finally, the last four columns represent the item category thresholds. From inspection, it would appear that these items appear to capture a wide range of the latent dimensions as defined by the eight variables.

**Table 3: Estimated item parameters for eight meta-perspective items (3 latent dimensions)**

<table>
<thead>
<tr>
<th>Item</th>
<th>f1</th>
<th>f2</th>
<th>f3</th>
<th>a (f1)</th>
<th>a (f2)</th>
<th>a (f3)</th>
<th>b1-2</th>
<th>b2-3</th>
<th>b3-4</th>
<th>b4-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.84</td>
<td>.09</td>
<td>-.03</td>
<td>1.52</td>
<td>.09</td>
<td>-.03</td>
<td>-1.06</td>
<td>.38</td>
<td>1.38</td>
<td>2.01</td>
</tr>
<tr>
<td>2</td>
<td>.91</td>
<td>.01</td>
<td>.03</td>
<td>2.15</td>
<td>.01</td>
<td>.03</td>
<td>-1.05</td>
<td>.30</td>
<td>1.37</td>
<td>1.95</td>
</tr>
<tr>
<td>3</td>
<td>.76</td>
<td>.03</td>
<td>.18</td>
<td>1.18</td>
<td>.03</td>
<td>.18</td>
<td>-.89</td>
<td>.72</td>
<td>1.49</td>
<td>1.98</td>
</tr>
<tr>
<td>4</td>
<td>.59</td>
<td>.05</td>
<td>.27</td>
<td>.73</td>
<td>.05</td>
<td>.28</td>
<td>-.72</td>
<td>.78</td>
<td>1.62</td>
<td>2.04</td>
</tr>
<tr>
<td>5</td>
<td>.04</td>
<td>-.03</td>
<td>.92</td>
<td>.04</td>
<td>-.03</td>
<td>2.38</td>
<td>-.79</td>
<td>.65</td>
<td>1.50</td>
<td>2.06</td>
</tr>
<tr>
<td>6</td>
<td>.16</td>
<td>.11</td>
<td>.63</td>
<td>.16</td>
<td>.11</td>
<td>.81</td>
<td>-.97</td>
<td>.32</td>
<td>1.27</td>
<td>2.07</td>
</tr>
<tr>
<td>7</td>
<td>.12</td>
<td>.42</td>
<td>.30</td>
<td>.12</td>
<td>.46</td>
<td>.32</td>
<td>-.72</td>
<td>.54</td>
<td>1.49</td>
<td>1.99</td>
</tr>
<tr>
<td>8</td>
<td>.02</td>
<td>.94</td>
<td>.02</td>
<td>.03</td>
<td>.63</td>
<td>.02</td>
<td>-.84</td>
<td>.32</td>
<td>1.05</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Note: \( f^* \) = factor loading, \( a(f^*) \) = item discrimination parameter, \( b \) = item difficulty/threshold parameter. N = 1,255.

5. Discussion

In summary, statistical results indicate that as many as three dimensions may underlie the responses to the eight meta-perspective items. A substantive interpretation of the dimensions found to underlie the responses to the eight meta-perspective items clearly supported a three-dimensional structure. Dimension 1 (items 1 thru 4) refers to a general self-perception of what one’s supervisor thinks of one’s work ethic.
dimension is comprised of the extent to which one believes that his or her supervisor thinks one tries to “do things better at work,” “do more than asked,” “work hard,” and possesses confidence in one’s “ability to succeed, reach challenging goals, or overcome obstacles.” Dimension 2 (items 5 and 6) refers to a self-perception of supervisor’s thoughts on work products. Specifically, these items reflect a meta-perspective of the extent to which one completes work on time and produces high quality work. Lastly, Dimension 3 (items 7 and 8) reflects a self-perception of one’s supervisor’s thoughts on personal self-regulation. This dimension includes items measuring the meta-perception of work habits like “laziness, length of breaks, etc.” and the extent to which the subordinate thinks that the supervisor thinks he or she has to “check up on” the subordinate.

Based on examination of the factor loadings and item discrimination indices, it appears that the inclusion of item number 7, “I think that my supervisor thinks that my work habits (laziness, length of breaks, etc.) are excellent,” may need to be re-evaluated, as its potential double-barreled wording. Specifically, we recommend that changing the item to “I think that my supervisor thinks that my habits regarding how I structure my work are excellent” might garner better results, as “habits” are defined conceptually rather than through multiple examples. If modification does not improve its overall fit with Dimension 3, eliminating item 7 entirely would result in a two-dimension scale, which might be a more accurate representation of the scale. In that case, future researchers seeking to validate the scale also would have the opportunity to augment Dimension 3 with additional items as deemed necessary based on their results.

Three strengths of our approach increase the contribution of the present study to the management literature. First, this study represents an important first step in empirically testing the social relations model in the context of human resource management generally, and in performance appraisal specifically. Performance management is a perennially challenging area for practicing managers, and social relations theory has the potential to help us understand some of the miscommunications inherent in the process and address them proactively. However, the measurement challenges associated with a social exchange-based, reflexive theoretical perspective have not readily lent it to empirical study in the management literature. Scale development in this area is an important first step. Secondly, we gathered data from two organizations rather than limiting our data to a single source, thereby increasing the potential generalizability of our findings. Additionally, the samples were fairly large and representative of their respective organizations. Finally, our use of multidimensional item response theory may help to increase awareness of this methodological approach in the management literature. MIRT methods allow researchers to simultaneously model the data and item-level measurement characteristics, thereby combining the strengths of factor analysis and item response theory.

6. Directions for future research

Despite the strengths of this study, however, three steps for continued research in this area appear warranted. First, although the use of two data sources may represent a good start, further replication is needed to determine how the findings reported here correspond to the results of studies conducted in other work environments. Following replication, a logical next step would be to examine the construct validity of the meta-perspective scale through convergent and discriminant validity strategies (Hinkin, 1998). Finally, once construct validity of the scale has been established, work will be needed to derive and test hypotheses regarding the antecedents and outcomes of the accuracy of meta-perspectives based on the social relations model. To date, this work has not been done due to difficulties operationalizing meta-perspective accuracy. Based on our scale development, we suggest that this hypothesis testing may best proceed by employing an analysis of self-other agreement on items in the meta-perspective scale.

Based on prior theoretical work in this area, we believe that empirically testing performance-based models that incorporate a meta-perspective may provide researchers with several important insights. As discussed, antecedents to a meta-perspective are thought to be primarily individually based, including variables such as self-identity, individual behaviour, and attributions. Therefore, empirical tests may help us to proactively recognize and address potential performance troubles before they escalate.

We also believe that variance in several work-related outcomes, including those examined under the rubrics of leader-member exchange theory, organizational justice, organizational support, and organizational politics will be accounted for by the meta-perspective construct given the complex exchange-based nature of
these research areas. Additionally, this research suggests that developing additional meta-perspective scales related to other organizational phenomena such as change management, organizational learning, and negotiation may be a useful endeavour. Given that these work-related outcomes may include group and even organizational level effects, meta-perspectives also may prove to be an important link between micro and macro levels of analysis, thereby providing future researchers with an important avenue through which meso-level research may be pursued (Porter, 1996). An additional opportunity to apply the meta-perspective in the context of performance appraisals may be in upward feedback (i.e., 360-degree) research. Identifying how managers see others as seeing their performance may even be more useful than identifying how managers see their own performance. Such insights may have significant applications for leadership research.

Appendix

Meta-Perspective Items
1. I think that my supervisor thinks that I always try to do things better at work.
2. I think that my supervisor thinks that I always try to do more than what's asked.
3. I think that my supervisor thinks that I always try to work hard.
4. I think that my supervisor thinks that I have a sense of confidence in my ability to succeed, reach challenging goals, or overcome obstacles.
5. I think that my supervisor thinks that I always get things done on time.
6. I think that my supervisor thinks that I never disappointed him or her with the quality of work that he or she receives from me.
7. I think that my supervisor thinks that my work habits (tardiness, length of breaks, etc.) are excellent.
8. I think that my supervisor thinks that he or she never has to check up on me.

References


Proposals for Designing and Controlling a Doctoral Research Project in Management Sciences

Jacques Lauriol
Groupe ESC Rouen, Mont-Saint Aignan Cedex, France

Abstract: Analysis of available literature on the design and control of a doctoral research project has led to the updating of a certain number of “constants” concerning the difficulties PhD students encounter during this process. These “constants” revolve around the crucial yet arduous issue of what comprises a “good thesis”. The present article offers a response to this question. It begins by highlighting the characteristics of the institutional context in which a doctoral study is carried out before offering three topics for thought (definition of the research project, deployment and evaluation of results). It concludes by focusing on the tension underlying all scientific research between “dissidence” (i.e., the ability to question “normal science”) and “conformity” (the study’s fit in an institutional context).

Keywords: Designing and controlling a research project, doctoral research in strategic management.

1. Introduction

Conducting and piloting doctoral research in Management, and maybe especially in the field of management (due to the porosity of this concept and its fundamentally paradoxical nature, de Woot, 1995 - Lauriol, 1999 – Dery, 2001 – Greenwood et al., 2004) is tantamount to “managing” the dynamics of an intellectual journey that is at best heuristic, and which more often than not involves an exploratory and even chaotic type of learning (Van de Ven, 1999).

This journey, which targets the fulfilment of a knowledge project concerned with an object of research, fits into a particular context, one defined by the rules, procedures, principles, paradigms and conventions that codify the conditions under which scientific knowledge is being produced in a given discipline (with knowledge being construed here as an “Institution”, Cheng et al, 1996). It is no easy thing to fit into this context, since the difficulties punctuating the insertion process are subject to debates and controversies that can at times become somewhat acerbic. Moreover, they feature a “situated” aspect, meaning that they are tied to the places where the research is actually being conducted (the Supervisory Laboratory, the Research Supervisor, the jury members, real activities out in the field, etc.).

Such debates can be organised into questions and answers that the present article will attempt to analyse. Our ambition is to establish a (necessarily partial) synthesis of these lines of questioning by situating them, first of all, in a four-dimensional space, one we deem capable of characterising the institutional context of research in Management today (Section 1). Section 2 will deal with those issues we consider to be the most important, and will try to offer a few elements of response. Three main themes will be developed: that of the definition of the research project in relation with established knowledge, that of the implementation of the project itself and of the mastery of apparently “chaotic” dynamics thrown up by this implementation. Finally, there is the delicate matter of the evaluation of a completed Doctoral project. All of this should enable us to conclude with a personal assessment of the difficulties awaiting those who try to design and pilot a doctoral research project. These include the existence of a sort of “dilemma” resulting from the tension between “dissidence” (i.e., an attitude of critical distanciation required for an assessment that will be “correct” in “common sense” and “preconception” terms (Bachelard 1983), on one hand, and “conformity”, on the other (given that research is part of an institutional context that cannot be ignored). What this implies is an ability to fit in (or comply) with the rules or conventions characterising the context, ones that more or less stringently define the criteria for developing and assessing a “good thesis”.

2. An institutional context for doctoral research in management

This institutional context can be characterised by an architecture comprised of four poles (cf. figure 1). The first involves research question itself and the object that the study purports to investigate. The issue here is knowing what one is looking for, i.e., defining a research question that will allow the researcher to build and develop a knowledge project relating to the object under study. The formulation of this question is closely linked to disciplinary field of reference, and obviously to the researcher’s mastery of the knowledge characterising said field. This first pole is tightly
embedded into a whole set of elements pertaining to the epistemological and ontological stances that the researcher has taken (i.e., his/her intention). At an epistemological level, all research projects must fit into the major paradigms (naturalism, positivism, interpretativism, constructivism and postmodernism) upon which modern scientific research approaches are based (Lincoln and al. 2000, Wacheux, 1996). Each of these paradigms states an ontological hypothesis relating to the nature of the reality that it would like to study, and regarding the conditions in which the outcome knowledge is produced and validated.

In addition to a simple “conformity check” approach featuring these established paradigms, the “choice” of a positioning drives the researcher’s own ontology (Campbell, 1988). Researchers make their determinations in light of their own beliefs, representations and existential positions, all of which allow them to formulate a research intention, as well as the finality of their study (see the second pole). By so doing, they can offer a “justification context” (Boltanski, Chiapello, 1999) to legitimise their study within the disciplinary field to which it belongs.

Piloting a research project is therefore, at the very least, tantamount to producing an “optimal” equilibrium between these three poles. This dynamic equilibrium, which is probably relatively close to the punctuated equilibrium that M. Tushman and E. Romanelli have proposed (1994) determines to a large extent the fourth pole of this architecture, which concerns the relevancy of the project and the coherency of its outcomes.

The relevancy of a project can be assessed in two main dimensions:

- Its legitimacy in terms of the paradigms, schools of thought or research programmes that characterise a given disciplinary field.
- Its legitimacy in terms of a “social demand”. This results from an equilibrium between expectations that are academic in nature, as they are expressed in a disciplinary field at a given moment in time, and expectations of a world that can be qualified as “professional”, being the world of actors engaged in real activities (the world of practices and practitioners being affected by this research project). These expectations are not necessarily convergent and can pursue logics that are complementary yet also concurrent and even antagonistic.

Coherency can be analysed in terms of the research project’s architecture (or design), which in turn conditions the validity and the reliability of its outcomes. The articulation between a research approach and literary analysis, the validity of the proposed conceptual construct, the methods applied when gathering and processing the data used plus the analysis and interpretation of the research outcomes are all part of this coherency.

The study’s assessment necessarily fits within a framework that is defined by a host of controversies concerning the conditions under which scientific knowledge is being produced. These controversies express strong oppositions between positions that are more or less contrasting (positivism versus constructivism, contents versus process, etc.). In this way, they ask questions about two essential dimensions of the knowledge production process: the causality established between the observations made and the analysis offered thereof; and its objectivity, which can be based on recognised protocols (explanation of facts depending on criteria of verifiability, confirmability and refutability). To design, organise and pilot a doctoral research project, it therefore seems crucial that the proponent seeks (and finds) a “satisfactory” positioning within this context and the different constraints that characterise it. It is in the search for this point of equilibrium that the researcher faces the delicate problem of determining a

---

**Figure 1:** The institutional architecture of a research project in the field of management

The third pole combines all of the questions relating to “how?” What this involves is thinking about the methodological research mechanisms that are going to determine the validity and reliability of the knowledge being generated by the development of a research project. The choices made here, in terms of data production, compilation, processing and analysis methods will express or reflect the epistemological stance that the researcher has more or less explicitly assumed. They help to deploy paths that can be used to explore the object (be it theoretical, empirical or a construct) and enable the modes of (deductive, inductive and/or abductive) reasoning that will be mobilised to achieve the study’s objectives (KOENIG 1993).
position along a sort of continuum that is comprised of the two aforementioned extremes: dissidence on one hand; and the ability to manage “pressure towards institutional conformity” on the other. To cope with this problem, we suggest three dimensions that appear to be crucial for the design and conduct of a research project.

3. Three questions and a few answers regarding the definition and assessment of a ‘good thesis’

These three questions relate to three key steps in the design and completion of a research project:

- Definition of the project itself, based on a researcher’s “footing” in the paradigms or schools of thought that specify established knowledge in a discipline.
- Project Deployment, i.e., the researcher’s ability to manage a dynamic process necessitating the establishment of specific support modalities and mechanisms.
- Evaluation of the project once completed, meaning the delicate question of assessing whatever work has been done.

3.1 How to define a project in relation with established knowledge

This question, which relates to the researcher’s positioning in terms of a certain number of paradigms considered to be more or less established (or, to the contrary, emergent), can cause many problems. First of all, which criteria help us to define a paradigm? When can we consider that a paradigm is established and has become part of a “normal science”? Inversely, when must it be treated for the “anomalies” that characterise a crisis period, one that can be the precursor of a “revolution” and may bear an emergent paradigm (Kuhn, 1983)? Moreover, many Management Science disciplines are characterised by their great openness to a number of ancillary disciplines (economics and its many branches, psychology, sociology, etc.). This situation creates an additional problem, which is the project’s footing in a mono-disciplinary approach, or inversely in an inter- or multi-disciplinary one.

Of course, it is not easy to answer these questions. First and foremost, they depend on the researcher, and on the options s/he selects when defining and completing the study. They may also be a sort of “tactic” involving an adaptation of the disciplinary context to the project (or the opposite, with an adaptation of the project to a disciplinary context). Lastly, they are often initially codified within a given research structure (a laboratory or research centre), meaning that the researcher fits into an already established theoretical and methodological corpus. As an individual, the researcher must be “clear” about his/her ontological orientations. That is to say, s/he must be able to formulate the beliefs, representations or postulates underlying these orientations. This clarification is indispensable since it makes it possible to understand and justify the epistemological anchoring chosen to design and develop his/her study.

In other words, these are beliefs, and it is by clarifying them that one’s epistemological positioning can be justified and argued. For example, taking a constructivist position presupposes that the three postulates proposed by A. Giddens (1987) be “internalised”, or at least considered by the researcher. This is due to the need to be able to discuss the role that is delegated to a given agent or actor (“the return of the subject… and his/her subjectivity” Giddens .14) during the course of a collective action. It also means accepting and “recognising the fundamental character of agents’ practical conscience … their competency… their reflective capacities…” (Ibid. 33.41) whilst underlining “the contextuality of the social interaction… the diversity of the action contexts” (Ibid. 35.74) as well as the social systems’ production and reproduction function, which is driven by this interaction. This “self-aware” choice defines an actor (a manager, for example) as something more than a mere processor of information, an agent simply reacting to changing conditions. More than anything else, s/he is “an active participant in the construction of his own environment” (MIR et al., 2000.945).

This preamble to the formalisation of a research project often results from an interactive approach between the researcher’s lines of questioning: with knowledge established through literature, and with the paradigms (disciplinary ones, this time) that characterise a discipline. Here we find ourselves in a dual situation characterised by two poles: interdisciplinarity; and multidisciplinarity. This type of dilemma is frequent in the field of Management Sciences. Just consider the diversity of the theoretical approaches that become possible in scientific experiments when one focuses on questions like organisational revitalisation, change or transformation. The same applies to managerial and organisational cognition, and to Resource-Based strategic theories (Koenig, 1999). This diversity can induce the researcher to opt for a particular approach (a Resource-Based View instead of one formulated in terms of Dynamic Capacities), to seek complementarities, or to compare each of these
approaches so as to test their explanatory, predictive or operative potential.

This insertion in a approach that has been codified by a support structure mechanism which benefits from accumulated experience partially resolves the issue of paradigmatic choice. But only partially, since it does not dispel the problem of the researcher’s ability to work within the framework of these mechanisms and theoretical and methodological orientations, nor his/her ability to accept the “consequences” (i.e., participating advantageously in the controversies that derive from this positioning, supporting positions that can be relatively tenuous in terms of the “normal science” of a given era, etc.) In general, what is important with regards to this question of paradigmatic positioning is to avoid “caricatural antinomies” between different perspectives (Tsoukas 2000). These lead to the formulation of overly reductionist alternatives when the point is to seek complementarities and factors of coherency that make it possible to design and deploy a knowledge project that can be relevant (David, 2000).

Ideally, this positioning should help to ensure the development of a process that is necessarily iterative, non-linear and exploratory. Such a process targets stability, to be developed throughout this phase through the building of an internal coherency between the research study’s design, the issues it raises, the analytical framework that has been chosen and the outcomes. This brings us to the second approach under consideration, which is the management of time of a research project’s deployment dynamic.

3.2 Deploying a research project

This second dimension mainly concerns the problem of the study’s potential “feasibility”. It refers to the competencies mobilised by the researcher but also to the nature of the project itself (is it “actionable”?), its target audiences and the various support mechanisms at the researcher’s disposal when s/he is deploying the project. A research project’s feasibility dimension first raises questions about the competencies that the researcher is able to mobilise when developing the study’s various phases. This refers to his/her ability to acquire and to leverage assets (i.e., the knowledge that specifies a given disciplinary field), and to “coordinate them durably … so as to achieve his/her objectives” (Sanchez 2000.66)

It also refers to the theoretical understanding, which the researcher has of the disciplinary field in which s/he is operating. What are the main controversies, which (sometimes “implicitly normative”) postulates are they based upon, which theoretical or practical issues are at stake, to what extent is this of interest to today’s scientific community, and what is the “life expectancy” or strategic horizon of these controversies? (Weick 1999). These assets should enable the development of a cognitive maturation process revolving around the “three phases that structure a research process: the intuitive; the comprehensive; and the demonstrative” (Wacheux, 1996.2). Intuition results from a process of perceiving observed facts. Through analysis and reasoning (based on an epistemological anchoring), comprehension enables the formation of an initial yet “comprehensive problemation” of the research question being studied. This “problemation” should be questioned by testing facts relating to the initial representations that the researcher has developed, leading in the end to “reasoned intuitions” (ibid.7) that are the very basis of the “progressive construction of an ultimate explanatory proposal” (p.8).

When approaching a particular field of study, such analytical capacities should be linked to relational ones. This notion is strongly correlated with the project’s epistemological positioning. A “positivist” orientation necessitates a great deal of neutrality and a position of maximal externalisation. A constructivist orientation, based on the “acceptance of a universe built with actors’ representations” (de Bruyne 1974.13), mobilises observational qualities and aptitudes, as well as the ability to listen and to ask questions – all of which will allow the researcher to wield a certain control over any cognitive biases by which s/he may be motivated. These elements also raise questions of access to the field of real activity. This dimension is linked to the nature of the object and to the question of research. Can the relevant facts be modelled at what we can consider to be a reasonable level of complexity (multiplicity and diversity of variables)? Can they be manipulated out in the real world, and are there situations where this can be envisaged? Are they acceptable for a group of actors or for a target organisation, and what interests might they have in this process?

Aspects of this conjunction of interests between researchers and given fields of real activity introduce the notion of “social demand” into research. This concept can be expressed at two levels: the suitability of academic types of demands, based on an increase in knowledge in a given disciplinary field; and the expectations of the “pragmatic” world (Dery 1989), to wit, of a “client”
(the host organisation, which generally demands the resolution of concrete and practical problems). These expectations are not necessarily convergent or synchronous, and therein lies the rub. The problem’s treatment clearly depends on the nature of the study in question, and on its finalities (i.e., are we dealing with a realistic finality that seeks to describe the world such as it is, or an “instrumentalist” approach that tries to develop predictively-oriented theoretical mechanisms to enhance our understanding of reality).

In terms of this “social demand”, one should remember the distinction proposed by M. Latour (1987) and B. Callon (1989) between science and research. Research involves a science “that is in the process of being built” and which is based on a “scientific practice”, whereas science is “built” out of facts that have been established and validated through epistemological protocols. As a scientific practice, research cannot be conceived of outside of the context in which it develops (a mechanism or a network of relations). As a result, a “scientific” research project can only exist when it can be of “interest” to strategic actors involved in its network. In other words, “social demand” mainly results from the researcher’s ability to build a mechanism that mobilises these different actors. This brings us to the question of the researcher’s competencies - but also to the support mechanism where such competencies can be built and deployed. Indeed, it is within these systems (i.e.: Supervisory Laboratory) that a project takes shape and develops. These are also the places where we can build a “general agreement of minds” concerning the responses to be provided for the various points and issues raised previously.

This context can (sometimes dramatically) constrain the knowledge production process that brings the researcher to a certain intelligence of the object with which s/he is concerned. As such, and whenever possible, a study’s context must be chosen with care. This choice is a difficult one, due to the diversity of existing framework modalities and mechanisms. Such diversity reflects the varied nature of Research Supervisors, whose epistemological and theoretical stances towards one and the same problem can be quite diverse. This situation can be seen as something beneficial, due to the potential richness it creates in terms of the approaches’ (and therefore the projects’) diversity. But it is also risky, since it expresses the state of controversies characterising a disciplinary field at a given moment in time. To cope with these problems, the researcher must develop an ability to assess the general context so as to identify its “local rationality”. This assessment begins with the choice of a Research Supervisor, a decision that depends on a number of criteria (Xuereb 1999): his/her intellectual proximity to the project in question, academic reputation, availability, working methods and experience in supervising a thesis - not to forget the “relational proximity” as perceived by the Ph.D candidate. This dimension means that trust (a system of mutual expectations that is more or less calculated in nature) becomes the main mode for governing such a relationship.

Although the Research Supervisor constitutes the key element in this support mechanism, we should not neglect the resources that any particular structure has at its disposal (material or academic resources, membership in networks, etc.). This “stock” of resources constitutes the foundation of an “economy of accumulation” (resources that are complementary and additive, Cool 2000) and they cannot help but be beneficial to the Ph.D. research project’s development dynamics. Evaluations of a project’s feasibility are therefore situated in a three-dimensional space: the competencies that the researcher can mobilise; the subject’s relevancy; and its ability to produce results that can be activated at an academic and pragmatic level. There is also the question of access to a mechanism that will provide sufficient support for a research project’s deployment dynamics. It is within this space that a project can be designed and shaped in a way that leads to its being seen as a “good thesis”.

### 3.3 A few criteria for assessing a “good thesis”: the question of evaluation

Issues like what constitutes a “good thesis” or appropriate assessment criteria express the difficulties that young Ph.D. candidates face in developing a representation of what is expected of doctoral research. The problem is the same for Research Supervisors who have to codify requirements to frame an approach that fundamentally remains a process of (exploratory and heuristic) discovery. Thus, the response that is generally provided to the question “what constitutes a good thesis?” is that “this depends” - mainly on the subject and on the researcher. A “good thesis” is to a large extent a subject that is considered to be “interesting” or relevant in terms of the questions and controversies characterising its disciplinary field, the interests of the Research Supervisor and the “social demand” (the “pragmatic world” or societal subjects). A “good thesis” is obviously one that has been completed, i.e. one where a successful viva voce has been given. This is a dimension that could appear trivial but it is seemingly not very easy to “complete” a thesis. After all, this is a process whose ambition...
is to enrich the world of knowledge, and bringing a study to conclusion despite all the difficulties encountered requires confirmed competencies, time and a degree of tenacity. What this means is that a doctoral student has to learn to separate him/herself from any ambition that can be described as “excessive”, at least temporarily, by accepting the intermediate stage that is the viva voce. Doctoral research is training in research through research. Like all learning processes, one needs time to get close to the truth (“truth is the daughter of time” Pettigrew 1990. 271). Doctoral research constitutes the first step of this process, one whose objective consists of enabling learning, and of evaluating outcomes. This means having the ability to prepare oneself for the aforementioned separation, progressively detaching oneself from any overly precocious ambitions in order to avoid falling prey to “the melancholy” of a “lost object” (Freud, 1915).

A “good thesis” is also a research study that supports a thesis, insofar as it takes a position. When the project is underpinned by a subject and by an approach, the researcher has taken a position towards the debates and controversies relating to this approach. This implies a clear presentation of such tensions and means that it should be possible to justify the significance of the research questions being associated with them. “Taking a position” also means offering working hypotheses and a plan capable of covering the question under study. It also means the ability to offer precise responses, even if they are incomplete or full of ambiguities. This dimension raises the issue of the outcomes achieved within the study’s framework. Above and beyond purely methodological questions, outcomes mainly reflect epistemological positioning’s but they also involve presentation issues. A “good thesis” uses a high quality argumentative process to highlight the study and its outcomes. Here one considers the clarity of the overview and the rigorousness manifested in its manipulation of concepts - as well as the style of the narration, which should convince the reader of the authenticity, plausibility and critical nature of the work being done. Lastly, there is the ability to produce a text that is “limpid” at a semantic level, facilitating access to complex contents and allowing for an analysis that is “economic” at a cognitive level (Czarniawska, 1999).

In short, all of these argumentative processes must try to highlight the value of the thesis and its contributions, “defending” and affirming a “point of view” relating to a given problem whilst underlining the difficulties encountered; the study’s limitations, and the avenues of development it envisages. The conception and deployment of a research project thus appears as a singular approach with a permanent tension between “dissidence” and “conformity”.

4. A few elements of conclusion

Designing and conducting a doctoral research project: between Dissidence and Conformity? Designing and piloting a doctoral research project consists of managing a knowledge production process that revolves around three main sequences:

1. The building of whatever “Positioning” is needed to situate the project within a world of controversies, comprised of theoretical and epistemological stances that can be more or less in competition with one another. The idea here is to lay foundations or to outline a perspective based on an analysis of such controversies’ “social contents”, thereby allowing the researcher to assimilate this context, and to form an initial line of questioning.

2. Management of the research project deployment process. This second sequence is generally tied to the interaction between the researcher and the field of real activity. It often leads to new lines of questioning that can consolidate things or make them more fragile (during the study’s starting phase). The goal is to manage the evolutionary dynamic of a process that can be based on an incremental logic or on “punctuated equilibrium” (periods of continuity, incremental or fluctuating changes, sometimes leading to radical transformations (Romanelli et al. 1994).

The question here is what mechanism is in place to manage and regulate such tensions. The support system should be able to accommodate (i.e., by incorporating new schemas of thought and action into pre-existing schemas. Piaget. 1992) any new data considered to be problematic. It should also help to restore a renewed equilibrium, one that will be based on an “enrichment” of the researcher’s cognitive structure.

3. The “discursive structuring” of the project and its contributions to scientific knowledge. The point here is to render intelligible complex phenomena that are of interest to actors characterised by varying expectations. This involves a translation formulated by the key actor in this process (the researcher) who, after analysing the context and the legitimisation of his/her framework of analysis, offers a problematisation for the observed facts and “states” “intelligible links between heterogeneous activities (or facts)” (Callon. 1992.65).
In short, designing and piloting a research project is tantamount to one's ability to manage the dynamic process of building an equilibrium between two poles: "dissidence" on one hand, and "conformity" on the other. By equilibrium, we mean "a situation where individuals do the best they can for themselves given ... the institutional framework that defines the options they are being offered, and which interconnect their actions" (Kreps. 1990.6). "Dissidence" probably constitutes the main factor activating this tension, for as long as the project runs. This is an attitude or an intellectual stance that tries to ask questions about "normal science", whilst uncovering the sometimes implicitly normative theoretical postulates being conveyed by certain paradigms or schools of thought. It protects people from the "fads" that can characterise a theoretical context at a given moment in time and helps them to resist any "hegemonies", as suggested by B. Latour (1995), who considers that there is a need to "protect researchers and science itself from hegemony ... fads or domination" (81) but specifies "the need to be simultaneously independent and to know how to resist hegemonies", as well as the necessity of having "colleagues you can depend on" (ibid).

This stance implies the mobilisation of intellectual competencies that are based on an excellent knowledge of the disciplinary field concerned by the project. It also implies distanciation and the creative criticism capabilities that will make it possible to forge a project that is relevant and whose legitimacy can be established. The project legitimacy dimension can only be conceived of outside of an "agreement of minds", as expressed by an academic community (with research being construed here as an institution). A researcher cannot "be the only one to get it right" and preserve critical capabilities whilst affirming his/her thesis. What s/he can do is to act upon any social influence, thus guaranteeing the scientific nature of his/her approach and suggested conclusions. This harks back to the aforementioned dimensions (coherency and relevancy of the project, coherency of the "discursive structuring", quality of the argumentative processes and validity or plausibility of the outcomes, etc.). To these qualities, we can probably add the researcher's manifest ability to wield "a minority influence".

This ability, largely analysed in social psychology (Levine et al 1990), results from a style of behaviour that a reference group perceives as being consistent (i.e., constantly affirming a position, without contradictory statements, and based on the development of a system of proven logic, in reference to a logical framework of analysis). This ongoing consistency, when perceived by the group, means that the individual will be awarded a certain right to exert a social influence on the milieu in which s/he is acting, thus legitimising his/her study. It is probably in this capacity for minority influence that we find the key to building a dynamic equilibrium between "dissidence and conformity". It is a process that helps to "renovate" the study's context (all else remaining equal) by preserving the possible development of a (more or less dissident) "thesis", and by engaging a testing process that is in tune with established canons of epistemology and methodology (pressure towards conformity).

In lieu of a conclusion, we can probably consider that designing and piloting a doctoral research project involves a truly strategic approach (devising a position based on resources and competencies, as well as action plans that are meant to act upon an environment instead of accepting its constraints). In other words, this project of discovery involves a modicum of pleasure (like any intellectual journey), spiced by a host of difficulties (and pains). It is probably by situating oneself within such an orientation that the researcher can optimise his/her management of the separation dynamics to which, as aforementioned, s/he is subject.

References
Callon, M. 1989. La science et ses réseaux. La Découverte.


Latour, B. 1989. La science en action. La Découverte.


Validation of Simulation Based Models: A Theoretical Outlook

Morvin Savio Martis
Manipal Institute of Technology, India
oceanmartis@yahoo.com

Abstract: Validation is the most incomprehensible part of developing a model. Nevertheless, no model can be accepted unless it has passed the tests of validation, since the procedure of validation is vital to ascertain the credibility of the model. Validation procedures are usually framework based and dynamic, but a methodical procedure can be followed by a modeller (researcher) in order to authenticate the model. The paper starts with a discussion on the views and burning issues by various researchers on model validation and the foundational terminology involved. The paper later highlights on the methodology and the process of validation adopted. Reasons for the failure of the model have also been explored. The paper finally focuses on the widely approved validation schemes (both quantitative and qualitative) and techniques in practice, since no one test can determine the credibility and validity of a simulation model. Moreover, as the model passes more tests (both quantitative and qualitative) the confidence in the model increases correspondingly.

Keywords: Validation, simulation, dynamic models, validation schemes, validation process, modelling.

1. Introduction

Validation has been one of the unresolved problems of systems modelling (Mohapatra 1987). It is true for simulation models in general and system dynamic models in particular. System dynamic modelling makes use of computer simulation (packages like MataLab, Stella) to generate the consequences for studying the dynamic behaviour of the system. In contrast, validations of Optimisation Models, Decision Theory or Game Theory are often not questioned since the solution procedures are elegant and correct. Reasons for conceptual and simulation models having received more criticism could be the ease with which the models and their overall results being understandable. Another reason being, the simulation model of any system could only be an approximation of the actual system, no matter the amount of time spent on the model building. Hence if the model produced is not a ‘close’ enough approximation to this actual system, conclusions derived from such model are likely to be divergent and erroneous, leading to possible costly decision mistakes been made (Ijeoma et al. 2001).

According to Law (2001), validation can be done for all simulation models regardless of whether their corresponding systems exist presently or would be built in future. Also, Kleijnen (1999) and Sterman (1984) give insight on validation of simulation models using statistical techniques and reasoned that the technique applied would depend on the availability of data in the real system. Contradicting the above authors, some authors have also stated that “there is no such thing as an absolutely valid model, credibility of a model can be claimed only for the intended use of the model or simulation and for the prescribed conditions under which the model or simulation has been tested” (DMSO 1996). Sterman (2000) also argue “validation and verification are impossible; the emphasis should be more on model testing i.e. the process to build confidence that a model is appropriate for the purpose. Some models may be better than others; some models, while not completely valid, possess a greater degree of authenticity than others. Furthermore, all models are, in a sense, wrong because there could always be a counter test to which the model did not conform to completely”.

Nevertheless, the power of a model or modelling technique is a function of validity, credibility, and generality (Solberg 1992). Hence model validation is not an option but a necessity in a dynamic modelling scenario. Usually the simplest model, which expresses a valid relation, will be the most powerful; however, there is no single test that would allow the modellers to assert that their models have been validated. Rather, the level of confidence in the model can increase gradually as the model passes more tests (Forrester, and Senge 1980). The relationships of cost (a similar relationship holds for the amount of time) of performing model validation and the value of a model to the user as a function of model confidence are shown in Figure 1. As shown in the figure, the value of the model increases as the level of confidence in the model is increased, correspondingly the cost of model validation also increases.

Reference this paper as:
Validation cannot be carried out by the modeller (or researcher) alone; communication with the client (or user) plays a large role in building a valid model and establishing its credibility (Carson 1989). Another relevant issue of concern is that by how much the model output could deviate from system output and still remain valid (Kleindorfer et al. 1998). Since the model created is an approximation of the actual system, some errors and approximations are unavoidable. Model validation thus resides in decision between the modeller and client; when both groups are satisfied, the model is considered valid (Goldberg et al. 1990).

A wide range of tests to build confidence in a model have been developed by authors like Forrester and Senge (1980), Barlas (1989 and 1996), Khazanchi (1996) and Saysel et al. (2004); a summary of which is presented under Validation Schemes.

### 2. Validation defined

The definitions of validation as stated by different authors are listed below:

- Substantiation that a computerised model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model (Sargent 2003).
- Validation is the process of determining that the model on which the simulation is based is an acceptably accurate representation of reality (Giannanasi et al. 2001).
- Validation is the process of establishing confidence in the usefulness of a model (Coyle 1977).
- The process of determining the degree to which a model is an accurate representation of the real-world from the perspective of the intended uses of the model (DoD 2002).

### 3. Viewpoints on validation

The viewpoints on validation are based on modified views of traditional validation techniques. These characteristics of validation are listed accordingly below:

- A model should be judged for its usefulness rather than its absolute validity.
- A model cannot have absolute validity but it should be valid for the purpose for which it is constructed.
- There can be no one test with which the model validity can be judged.
- As a model passes the various tests, confidence in the model is enhanced.
- "Failing a test helps to reject a wrong hypothesis, but passing is no guarantee that the model is valid" (Sushil 1993).
- "Quantitative as well as qualitative validity criterion should be given more credence (Forrester 1961)."
- Most of the information from the real system is used to check the consistency of model behaviour.
- Rejecting a model because it fails to reproduce an exact replica of past data is not acceptable.
- Rejecting a model because it fails to predict a specific future event is not acceptable because social systems operate in wide noise frequencies.

### 4. Methodology for validation

Validation deals with the assessment of the comparison between ‘sufficiently accurate’ computational results from the simulation and the actual/ hypothetical data from the system. Validation does not specifically address how the simulation model can be changed to improve the agreement between the computational results and the actual data. The fundamental strategy of validation involves identification and quantification of the error and uncertainty in the conceptual/simulation models, quantification of the numerical error in the computational solution, estimation of the simulation uncertainty, and finally, comparison between the computational results and the actual data. Thus, accuracy is measured in relation to actual/hypothetical data, our best measure of reality. This strategy does not assume that the actual/hypothetical data are more accurate than the computational results. The strategy only asserts that simulation results are the most faithful reflections of reality for the purposes of validation (AIAA 1998).
5. Model validation process

Figure 2 shows the model validation process in a simpler form. The ‘problem entity’ is the system (real or proposed – e.g. dynamics of integrated Knowledge Management and Human Resource Management can be considered as a problem entity (Martis 2004)) to be modelled; the ‘conceptual model’ is the mathematical/verbal representation (influence diagram) of the problem entity developed for a particular study; and the ‘computerised model’ is the conceptual model implemented on a computer (simulation model). The inferences about the problem entity are obtained by conducting simulations on the computerised model in the experimentation phase.

Figure 2: Model validation process.

There are three steps in deciding if a simulation is an accurate representation of the actual system considered, namely, verification, validation and credibility (Garzia et al. 1990). ‘Conceptual model validation’ is the process of determining that the theories and assumptions underlying the conceptual model are correct and that the model representation of the problem entity is “reasonable” for the intended purpose of the model. ‘Computerised model verification’ is the process of determining that the model implementation accurately represents the developers’ conceptual description of the model and the solution to the model (AIAA 1998). ‘Operational validation’ is defined as determining that the model’s output behaviour has sufficient accuracy for the model’s intended purpose over the domain of the model’s intended applicability (Sargent 2003). Operational validity determines the models credibility. ‘Data validity’ is defined as ensuring that the data necessary for model building, model evaluation and conducting the model experiments to solve the problem are adequate and correct (Love et al. 2000).

6. Reasons for failure of models

Some of the reasons due to which the models fail the validation tests are enumerated below as follows:

- Model-structure- In both the conceptual model and the simulation model mathematical simplifications might be inadequate for capturing complex dynamics.
- Numerical solution- The solution of the simulation model might differ dramatically from the ideal solution.
- Input values- Proper numerical values of the inputs that describe the scenario for prediction might be known only approximately.
- Observation errors- Inaccurate observations of real system.
- System noise- Failure to recognize random changes existent in the system.
- Project management errors- These errors revolve around project management and related communication issues (Carson 2002).
- Inappropriate simulation software – either too inflexible or too difficult to use (Law 2003).
- Misinterpretation of simulation results

7. Validation schemes

7.1 Validation scheme as proposed by Forrester and Senge (1980):

This validation criterion is used to validate quantitative as well as qualitative models. The validation scheme is mainly divided into four phases; as the model passes more tests under every phase the confidence in the model increases correspondingly. The validation scheme as proposed by Forrester and Senge (1980) is enumerated below:

7.1.1 Importance of model objective:

The validity of a model cannot be greater than the objective set for it. Therefore, the model objective must be a justified representation of the values prevalent in the real system. The method of setting model objectives by the conceptualization of problems in the existent system seems unstructured, unless the problem elicitation is done under the guidance of experts from the various subsystems existent within the system. A model could be proven valid by a series of methods, but the validation may be totally useless if the objectives are wrongly set.

7.1.2 Validating model structure:

These tests help in establishing confidence in the model structure.
Tests of suitability:

Structure-Verification Test: This test is meant to answer the following question “Is the model structure not in contradiction to the knowledge about the structure of the real system, and have the most relevant structures of the real system been modelled?”

Dimensional-Consistency Test: “Do the dimensions of the variables in every equation balance on each side of the equation?” This test verifies whether all equations are dimensionally constant.

Extreme-Conditions Test: “Does every equation in the model make sense even if subjected to extreme but possible values of variables?” Policy equations are scrutinized for their applicability in extreme conditions.

Boundary-Adequacy Test: This test verifies whether the model structure is appropriate for the model purpose (Barlas 1989). “Is the model aggregation appropriate and includes all relevant structure containing the variables and feedback effects necessary to address the problem and suit the purposes of the study?”

Tests of consistency:

Face validity test: “Does the model structure looks like the real system? Is it a recognisable representation of the real system? Does a reasonable fit exist between the feedback structure of the model and the essential characteristics of the real system?”

Parameter-Verification Test: Parameters and their numerical values should have real system equivalents. “Do the parameters correspond conceptually and numerically to real life? Are the parameters recognisable in term of real systems, or are some parameters contrived to balance the equations? If the values selected for the parameters consistent with the test information available about the real system?”

Test of utility and effectiveness:

Appropriateness for audience: “Is the size of the model, its simplicity or complexity, and its level of aggregation or richness of detail appropriate for the audience for the study?” The more the appropriate a model for the audience the more will be the audience’s perception of model validity.

7.1.3 Validating model behaviour:

These tests help in establishing confidence in the model behaviour.

Tests of suitability:

Parameter sensitivity test: “Is the behaviour of the model sensitive to reasonable variations in parameter values, i.e. do the modes of the behaviour change with parameter variations?”

Structural sensitivity test: “Is the behaviour of the model sensitive to reasonable structural reformulation, i.e. do the modes of the behaviour change with structural variations?”

Tests of consistency:

Behaviour-Reproduction Test: Here the generated model behaviour is judged with the historical behaviour. “How well the model generated behaviour matches observed behaviour of the real system in terms of symptomatic generation, frequency generation, relative phasing, multiple mode, and behaviour characteristics?”

Behaviour-Prediction Test: This test calls for pattern prediction. “Whether or not a model generates qualitatively current patterns of future patterns of future behaviour in terms of periods, shape or other characteristics?”

Behaviour-Anomaly Test: Behaviour conflicting with the real system helps in finding obvious flaws in the model. “Does behaviour shown by the model is conflicting with the real system behaviour and how implausible behaviour arises if the assumptions are altered?”

Family member test: Whenever possible, attempt should be made to build a general model of the class of system to which a particular member belongs. The general theory is depicted in the structure. Parameter values are chosen to depict a particular situation. By choosing a different set of parameter values the model can be applied to other situation as well.

Surprising behaviour test: “Does the model under some test circumstances produces dramatically unexpected or surprise behaviour, not observed in the real system? Whether such a surprise behaviour is due to model structure or some causes in the real system can be assigned to such a behaviour?”

Extreme-Policy Test: “Does the model behave in an expected fashion under extreme policies, even ones that have never been observed in the real system?” If the model behaves in an expected fashion under extreme policies, then it boosts confidence in the model (Saysel et al. 2004).

Boundary adequacy (behaviour) test: “Does the model include the structures necessary to address the issues for which it is designed?” If an extra model structure does not change the behaviour, then this extra structure is not necessary. Alternatively, if a model structure does not reproduce desired model behaviour, it calls for inclusion of additional model structure (Barlas 1996).

Behaviour-Sensitivity Test: “Whether plausible shifts in parameters can cause model to fail
behaviour tests previously passed?” Here the sensitivity of the model to changes in parameter values is judged (Saysel et al. 2004).

**Statistical tests:** “Does the model behaviour statistically like data from real system?” (Law and Kelton 2000).

Tests of utility and effectiveness:

**Counter intuitive behaviour:** “In response to some policies, does the model exhibit behaviour that at first contradicts intuitions and later, with the aid of the model, is seen as a clear implication of the structure of the system?” (Richardson et al. 1981).

“Is the model capable of generating new insights or at least the feeling of new insights, about the nature of the problem addressed and the system within it arises?”

**7.1.4 Validating policy implications:**

Tests of suitability:

**Policy sensitivity and robustness test:** The sensitivity of a policy with respect to change in parameter values is judged during this test. “Whether the model based policy recommendations change with reasonable changes in parameter values or reasonable alteration in equation formulations?”

**Tests of consistency:**

**Changed Behaviour Prediction Test:** “Whether the model correctly predicts how behaviour of the system will change if a governing policy is changed?”

**Boundary adequacy (policy) test:** “Whether modifying the model boundary (i.e. conceptualisation of additional structure) would alter policy recommendations arrived by using the model?”

**System Improvement Test:** “Whether the policies found beneficial after working with a model, when implemented, also improve real system behaviour?”

Test of utility and effectiveness:

**Implementable policy test:** “Can those responsible for policy in the real system be convinced of the value of model-based policy recommendations? How is the real system likely to respond to the process of implementation?” the policy recommendations should be such formulated and argued so as to fit in the mental models of those to whom they are addressed.

**7.2 Validation scheme as proposed by Khazanchi (1996):**

This validation criterion is mainly used to validate qualitative/conceptual models and consists of a set of criteria for validation. The criteria for validation as suggested by Khazanchi (1996) are as follows:

1. **Is it plausible/ reasonable?** This criterion is useful to assess the apparent reasonableness of an idea and could be demonstrated by deduction from past research or theories.

2. **Is it feasible?** A feasible concept would be operational only if it would be open to graphical, mathematical, illustrative characterisation.

3. **Is it effective?** An effective conceptual model should have the potential of serving our scientific purposes (Kaplan 1964).

4. **Is it pragmatic?** This criterion emphasises that concepts and conceptual models should have some degree of logical self-consistency or coherence with other concepts and conceptual models in the discipline (Hunt. 1990).

5. **Is it empirical?** Empirical content implies that a concept or conceptual model must have “empirical testability” (Hunt 1990).

6. **Is it predictive?** A conceptual model that is predictive would, at the least, demonstrate that given certain antecedent conditions, the corresponding phenomenon was somehow expected to occur.

7. **Is it inter-subjectively certifiable?** This criterion states “Investigators with differing philosophical stance must be able to verify the imputed truth content of these concepts or conceptual structures through observation, logical evaluation, or experimentation (Hunt 1990).

8. **Is it inter-methodologically certifiable?** This criterion provides that investigators using different research methodologies must be able to test the veracity of the concept or conceptual model and predict the occurrence of the same phenomenon.

**8. Other validation techniques**

Combinations of these techniques are generally used for validating a simulation model. These tests can be used in addition to the validation schemes in the preceding section to increase the credibility of the model.

1. **Comparison to other models:** Different outputs of the simulation model being validated are compared to those of other ‘valid’ models.
2. Degenerate test: This has to do with appropriately selecting values of the input and internal parameters to test the degeneracy of the model’s behaviour. For instance, test to see if the average number in the queue of a single server continues to increase with respect to time when the arriving rate is larger than the service rate (Ijeoma et al. 2001).

3. Events validity: The events of occurrences of the simulation model are compared to those of the real system to see if they are similar e.g. verify the exit rate of employees.

4. Face validity: This has to do with asking knowledgeable people if the system model behaviour is reasonable (Forrester 1961).

5. Historical Data validation: The experimental data is compared with the historical data; to check whether if the model behaves in the same way the system does (Balci et al. 1982).

6. Predictive validation: The model is used to predict the system’s behaviour, and then comparison is made between the real system behaviour and the model’s forecast to determine if they are the same (Sargent 2003).

7. Schellenberger’s Criteria: This include technical validation which has to do with identifying all divergences between the model assumptions and perceived reality as well as the validity of the data used, operational validity which addresses the question of how important these divergences are and dynamic validation which ensures that the model will continue being valid during its lifetime (Ijeoma et al. 2001).

8. Scoring Model Approach: Scores (or weights) are determined subjectively when conducting various aspects of the validation process and then combined to determine category scores and an overall score for the simulation model. A simulation model is considered valid if its overall and category scores are greater than some passing score(s) (Gass 1993).

9. Clarity: Clarity refers to the extent to which the model clearly communicates the implied causality/linkages.

10. Black-box validation: This test is concerned with determining whether if the entire model is an adequately accurate representation of the real world (Ijeoma et al. 2001).

11. Extreme Condition Test: The model structure and output should be reasonable for any extreme and unlikely combination of values in the system. For example, if in-process inventories are zero, production output should be zero (Sargent 2003).

9. Conclusions

As rightly coined by (DMSO 1996), validation is both an art and a science, requiring creativity and insight. But validation is a convoluted, multifarious and exasperating procedure, and is unavoidable as it is the evidence for the steadfastness and legitimacy of the model. Moreover, no single procedure can suit all the models. Statistical based validation techniques have been widely accepted among the management community. But the problem associated with this method is being able to determine the suitable type of statistical procedure, which in turn depends on the right type of data that is available for analysis. Moreover, the amount of deviation from the real system that is within the acceptable limits is uncertain.

The paper has given an insight on the widely approved validation schemes and techniques in practice. The validation schemes can be applicable to quantitative (mathematical/computerised) as well as qualitative (conceptual) models. But reliability of the model can only be ascertained as the model passes more and more tests. Also, the decision of accepting a model as valid cannot be left to the modeller alone, inclusion of the client / practitioners in the validation procedure should be ascertained. Researchers and practitioners may find this paper quite useful as the procedures for validation discussed are quite generic, and hence, may be applied to other dynamic models as well.

References


Motivators for Australian Consumers to Search and Shop Online

Ian Michael
College of Business Sciences, Zayed University, Dubai, UAE
Ian.Michael@zu.ac.ae

Abstract: This paper investigates the factors that motivate Australian consumers to use the Internet to search and shop for products and services. A qualitative research method was used, utilizing a semi-structured, in-depth interview technique. Twenty in-depth interviews were conducted, using a snowball sampling method. The study found that there were six motivating factors that drew consumers to search and shop online for products and services. These factors include: convenience, saving time, cheaper prices along with the ability to compare prices, good place to shop for specialty and hard-to-access products, higher level of consumer control, and the ease of comparing products and services.

Keywords: Motivating factors; Internet; consumer behaviour; search; shop; online

1. Growth of the Internet

Worldwide estimates of electronic commerce activity in the future vary but they are all uniformly staggering; for example, the OECD’s studies suggest that sometime between 2003 and 2005 electronic commerce activities could grow and reach around $1 trillion (Prabhaker, 2000). Other studies indicate that approximately US$3.1 billion was spent online in 1999; this was estimated to reach $15 billion by 2003. Over its history, retail e-commerce revenues increased from $8 billion to $18.6 billion in 1999 and were predicted to reach $80 billion by 2003 (Joines, Scherer and Scheufele, 2003). Hence in this age where the Internet plays a major role in business activities, it is imperative for marketers to gain a thorough understanding of consumer behaviour in this online environment. In studies conducted by Forrester Research, online retailing in the USA reached $108 billion by the end of the year 2003. In the United Kingdom (UK) where online retailing has just 0.2% market share of all retailing in 2003, retailing forecasters have that this could multiply by as much as 15 times over the next three years (Karayanni, 2003). Park and Kim (2003) predict that between 2000 and 2005, the value of goods delivered to homes in the UK will rise by 83% from £18.9 billion pounds £34.5 billion, much of this growth fuelled by the steep increase in online shopping.

2. The Internet in Australia

The Internet is rapidly becoming a critical tool for marketers as consumers are being attracted to it as a place to search and shop for products and services. According to the Australian Bureau of Statistics (ABS), Australian household’s accessibility to the Internet at home grew from 16% in 1998 to 46% in 2002 (ABS, 2004). During the year 2003, the Internet was used by 23% of Australians to pay bills or transfer funds, an increase of 6% from 1998. The rate of shopping online in Australia increased too, with 15% percent of all adult Australians who shopped online in 2002, this was up from 11% in 2001 (ABS, 2004). Other studies conducted in Australia found that Internet access by the 18-24 year old group was as high as seventy percent. It has also been found that there seems to be a shift in the demographic make up of internet users, with ‘baby-boomers’ and senior citizens now using it more (Lunn, 2000). In a study conducted by the National Office of the Information Economy (NOIE) it was revealed that seventy five percent of Australians over the age of 16 years had access to the during the first quarter of 2003 (NOIE, 2003).

Lynch (2003) suggests that there may be a relationship between the types of products purchased online and consumer’s gender and age. According to the author studies conducted in the past have indicated that younger Australians preferred to buy music, while people over the age of fifty purchased hobby collections such as die cast cars. It was also reported that women in the age of fifty and above used the internet to search and shop for travel and tourism related products online. In a recent study jointly conducted by AMR and Forrester Research as reported by Lynch (2003), it was found Australians differ in their internet usage patterns compared to other places. Fifty eight percent of respondents said they went online at least once a month, four percent of these have broadband access at home, and fourteen percent say they were likely to get broadband in a years time. It was also found that Australians were motivated and lured to the Internet to search and shop for gadgets and entertainment purposes. Hence, most respondents agreed that they could
be classified as “gadget grabbers” and “media junkies”.

Studies pertaining to consumer behaviour over the Internet have been a neglected area in Australia this far. In a couple of very broad and basic studies done in the country, it has been found that Australians were motivated by the entertainment benefits that websites offer. The Australian Internet user was found to align strongly with statements such as “technology is important to me”, and “I like to impress people with my lifestyle”. Another study suggested, that more people used the Internet in the country because of falling costs of high-speed connections, this significantly increase the appeal of transacting over the Internet (Henderson, 2003). It was also found that Australians consider the Internet to be a risky medium when it came to making payments. In comparison, British Internet buyers were less worried about online payments. In Australia, only 20.6%, the lowest among the nations surveyed say they had bought online but paid offline. Foreshew (2003) forecasts that consumer e-commerce will grow in Australia from $4.8 billion in 2002 to about $18 billion in 2006. Longer working hours, more expensive petrol, and the desire to spend more quality time with friends and family will act as the key drivers for this e-tailing growth. According to Henderson (2003), the most popular consumer purchases in Australia are LCD Notebooks and digital cameras, these products account for more than fifty percent of sales. In their annual research report on e-commerce, the National Office of the Information Economy (NOIE) found that the most popular products purchased online in 2002 were as follows:

- Multi category computers and consumer electronics (33%)
- Special interest news (29%)
- Music (29%)
- Classified/Auctions (28%)
- Videos/Movies (25%)
- Shopping directories and guides (17%)
- Online gaming (15%), and
- Arts/graphics (14%)

(NOIE, 2003)

3. Consumer motivations and the Internet – studies from other countries

In the Internet shopping context, consumers expectations are rapidly changing, they expect better pricing, ease of use of websites, secure payments guarantees and a comprehensive customer service back up which they would normally receive from traditional channels of distribution. As the Internet is rapidly becoming a major source for the search and purchase of products and services, there is an urgent need from marketers to achieve the same levels of service that is provided in a traditional marketplace. There is a great demand for research to investigate consumers, and their buying behaviour over the Internet and other electronic channels.

It has been emphasised that research is yet to examine the underlying motivations of consumers when they using the Internet (Joines, Scherer, and Scheufele, 2003). This is certainly true in the case of specific consumer-related activities such as searching for information on products and services and shopping online. These authors also suggest that there appears to be a lack of published research identifying how motivations may vary across different website activities. It has been found that using the web for the consumption purposes is increasingly becoming a primary reason that is attracting people to the Internet (Lever, 2003). In this context there are numerous studies suggesting that not only is online shopping on the increase, but also that consumers are spending more time searching for information on products and services over the Internet. Hence investigating the underlying motivations and concerns of consumers’ who use the Internet to seek product and service information is a topic worthy of investigation. There is a demand for answers pertaining to the possible underlying reasons as to why people use the Internet especially when they use it to gain product knowledge and/or to shop for products and services. The answers to these important matters can also enable the advertising function in marketing to target audiences and tailor web content more effectively.

With regards to the theory of motivation, this paper addresses a key question pertaining to understanding the main motivators for consumers to visit and search websites and why they shop online. Researchers and authors in the past have suggested that further empirical research needs to be carried out to investigate motivators based on variables such as different segments in a marketplace. This paper aims at understanding the key motivators for consumers to search and shop online. Consumer behaviour research into online search behaviour found that the people visited Internet sites only to gather information (Evans, Wedande, Ralston and Hul, 2001). Further research into this aspect by the authors found that information gathering fell into two broad categories, namely: personal and business.
Personal information collection included items such as sport, films, games, hobbies, and chat, whereas business information collection included technical information, academic research, and seeking information for the company (Evans et al., 2001).

In a study to ascertain differences between web shoppers and non-shoppers, Karyanni (2003) found that the most powerful discriminant variable between web shoppers and non-shoppers was shopping motives. This was especially so when it came to issues pertaining to time efficiency, availability of 24 hours shopping and queues avoidance. Web shoppers seemed to be less concerned with the enjoyment derived from traditional shopping. This might be due to the limited hours that they have because of long work hours. In a study on the teenage segment, Grant and Waite (2003) found them motivated to spend time online because of an interest in entertainment and educational type products available over the Internet. Other research focused on the impact of customer convenience, and product type on shopping behaviour over the Internet. For example, Pazgal and Sikka (2001) hypothesised that for information based products like software and digital music, the Internet is good to serve all stages of the marketing process. It can simultaneously serve as a communication, transaction, and distribution channel thus creating vertically integrated channels. The author's suggest that Dell used this philosophy to integrate most of its business processes via the Internet. They (Dell) have now pursued an integrated marketing strategy to bring their online business model more inline with its “direct model” business process. Furthermore, Pazgal and Sikka’s (2001) study concludes that consumers purchased more software and digital products online because these products were easy to access online, and required no physical medium to transport them.

On the issue of convenience as, Lee and Marlowe (2003) stressed the need for marketers to understand the true meaning of ‘convenience’ when it comes to consumers using the Internet. They state that although most consumers view convenience as one of the most important decision-making criteria, their definitions of convenience vary across consumers. Though it was hard to generalise the findings from the various focus groups that the researchers used in their study, they found that “convenience” is defined differently depending on consumers’ age, life-style, and comfort level with electronic technology. Grant and Waite’s (2003) study on Internet usage among the youth market, found them to use the medium as a route to escape from boredom. It was also found that this segment used the Internet to establish and maintain communication with friends and family in order to maintain these social networks, and for academic purposes. Their study concluded that around 88% of those aged 16-24 has accessed the Internet at some time; this represented a significantly higher level than of any other age group. The authors highlighted the lack of empirical research into the adolescent and young adults’ segment.

In a study conducted to ascertain why consumers visit luxury product websites it was found that the most important reasons were: "to look for information" (40%), "leisure” (19%), and "to shop" (18%), other reasons were "to chat" and "email" (Riley and Lacroix, 2003). Jiang, Hsu, Klien and Lin (2000) found that longevity was an important driving factor for people’s utilisation of the Internet. It was noticed that the longer an individual has adopted the Internet the more likely he or she would continue to be there. Donthu and Garcia (1999) found that Internet shoppers were more convenience seekers, innovators, and less risk averse than non – Internet shoppers. Zhang (2001) suggested that online consumers have less discretionary time and were “time starved,” they were also found to be exhibiting a higher level of inertia in online stores in comparison to bricks-and-mortar stores. Rettie (2001) hypothesised that a key motivating factor for people to visit, search, and shop online was the consumer’s view that the Internet was a good place to provide information, access to explore and purchase “hard to find” goods.

Pazgal and Sikka (2001) are of the view that knowledge and a sense of power could be key motivating factors for consumers to constantly stay in touch with the Internet. According to them, consumers are aware that the more information they have, the more they can shift the balance of bargaining power. In their study, consumers stated that more information made them confident and sophisticated. The authors find this attainment of power to be a viable motivating force. As the Internet grows, consumers are becoming powerful in two ways: firstly by bonding together online as virtual communities and leverage their numbers to bargain, and secondly consumers find it easier to search for information about product attribute, prices, and possible substitutes. Hence the future marketplace will be a “pull based” one, in that demand for new products will be fuelled by consumer's needs rather than manufacturers suggestions (Pazgal and Sikka, 2001). In her research, Modhal (2000) found people used the Internet for varied purposes. One stated that he was a tin soldier collector and made use of the Internet to broaden his collection, another needed to learn about
household chemicals that helped with his job, and a third was organising her local senior centre social schedule online. The respondents highlighted three motivating factors for them to go online: career, family, and entertainment.

Teo (2001) found that a key extrinsic variable for usage was perceived usefulness; this played a greater role than perceived enjoyment in computer usage. It was also found that perceived usefulness affects intention to purchase while perceived ease of use does not. Earlier research has suggested that one motivating factor for consumers’ to search and shop for products online was consumers felt that the Internet offered lower prices. Rettie (2001) and Trochia and Janda (2003) disagreed with this finding by Teo (2001). These two authors found that online consumers were less concerned with lower prices, and more with favourable outcomes, and a satisfying shopping process. They concluded that low prices and quick delivery might not be enough to attract and retain customers. Sathye (1999) and Pazgal and Sikka (2001) suggest that as the Internet is a cheaper medium to transact and communicate consumers would flock to it. In one study it was found that Internet banking have significantly lower costs structures than traditional delivery channels, and could operate at an expense ratio of 15-20% compared to 50-60% for the average bank (Sathye, 1999). This proposition was reinforced by the Booz Allen and Hamilton’s study which found the following variable costs exist for banks at their transaction level: Teller $1.07, Telephone $0.52, ATM $0.27, PC Banking $0.015, Internet $0.01 (Pazgal and Sikka, 2001).

In a study on consumer behaviour over the Internet, Jarvenpaa and Todd (1997) found that consumers were motivated to use the Internet because it was a time saving device in shopping along, with reducing one’s shopping effort. Swaminathan, Lepkowska-White and Rao (1999) found that consumers who were motivated by convenience, they further stressed that people who value social interaction were less interested in using the Internet for shopping. These consumers shopped less over the Internet and generally spent less money on e-commerce.

In an evaluation of electronic service quality (e-SQ), it was found that perceived control over the shopping environment and perceived convenience were critical factors that drew people to the Internet (Zeithaml, Parasuraman and Malhotra, 2000). In a study pertaining to enquire how to convert searchers to shoppers online, Van den Poel and Leunis (1999) found that money-back guarantees, well-known brand names and price reduction techniques usually convert searchers into buyers. Furthermore it was found that a web’s reliability and fulfilment ratings had a positive effect on consumers’ intentions to purchase (Wolfinbarger and Gilly, 2002). It is imperative for Australian marketers to understand what motivates consumers to search the Internet and websites for product information, and then what makes them shop online. Understanding this will lead to better-constructed websites with the objective of attracting consumers attention.

As discussed above, studies have indicated that convenience is a key driving factor to search and shop for products online. Others suggest that the act of gathering information for the purpose of making informed choices, and feeling more in control of the purchase decision is another key factor. Price and variety have also been suggested as reasons for people to go online to shop. However as Grant and Waite (2003) have suggested more empirical research needs to be carried out to investigate motivators. Consumer segmentation variables such as demographics, that include age, gender and occupation, need to be taken into consideration when researching motivating factors and the usage of the Internet.

4. Research method

This paper investigates key themes in online consumer behaviour. Its aims are to:

- Understand what motivates Australian consumers to search and shop for products and services over the Internet

This research incorporated for the study was a qualitative approach. This research approach is more intrusive and less structured than quantitative research techniques. It has been suggested that qualitative methods are appropriate when the research is exploratory in nature, and when the area for examination is unfamiliar to the researcher; this is the case with online consumer behaviour in Australia. The primary aim of the research was to gain an insight into what motivated Australian consumers to search and shop for products and services over the Internet.

According to (Ruyter and Scholl, 1998), qualitative research provides an in-depth insight; it is flexible, small-scale and exploratory, the results obtained are concrete, real-life like and full of ideas. The above two author’s further highlight the value this type of research, suggesting it can prove to be very useful for market research purposes especially in practice. They found it widely used by management consultants and public policy makers, where answers to questions in consumer decision-making such as why do consumers buy this product and not the other are important.
Ueltzhöffer and Ascheberg (1999) suggest that the principal aim behind qualitative methods in sociological and psychological market research is to elucidate the individual's subjective interpretative patterns, his/her personal experiences, viewpoints, motives- and thus his/her creative potential as well. Ruyter and Scholl (1998) stress the need to use qualitative methods especially to explore substantive areas about which little is known. This is the case in Australia, regarding the knowledge we have to date about motivations and consumer behaviour in an online environment. In addition, qualitative methods are extremely useful in obtaining the intricate details about phenomena such as feelings, thought processes, and emotions that are difficult to extract or learn about through more conventional research methods (Strauss and Corbin, 1998).

After investigating various models and approaches to qualitative research methods, in this study it was decided to take a research stand largely based on Patton's (2002) “reality testing” a positivist and realist type approach. Tesch’s (1990) method that propagates “the discovery of regularities approach” using an interpretive approach was also utilised. The final outcome of this study it was decided to take a research stand largely based on Patton's (2002) “reality testing” a positivist and realist type approach. As part of reality testing, a researcher needs to have a thorough understanding of the past theoretical issues through the literature. Reid (1995) suggests that in order to propose theory creation, the researcher needs to have a grasp of the relevant theory as well as knowledge of contemporary issues. Being a qualitative study, this research is particularly oriented towards exploration, discovery, and inductive logic. The themes, categories or dimensions of analysis emerge from open-ended interviews. The study utilised an inductive analysis; patterns, themes, and categories of analysis come from the data; they emerge out of the data rather than being decided prior to data collection and analysis.

The background literature that was collected on motivating factors in other parts of the world played a key role in creating the constructs for this study. Pole and Burgess (2002) highlight the role of substantive literature and its usefulness in elaborating the analytic framework. According to the authors the analyst needs a wider perspective on the topic. At the start of the study the formulated problem statement needs to be created on the basis of a rough overview of the field under study. As the framework for the study, this research utilised prior theory as its foundation.

5. Sampling technique

The study used the ‘snowballing’ or ‘referral’ or ‘chain’ sampling technique, where participants were requested to provide the researcher with at least one other person who could be contacted as a participant for this study. It was made clear that the person being referred would also need to be a researcher or shopper for products and services online. After speaking to them, the participant usually phoned back the researcher after a few days with a contact name, or would give these details during the interview, or in some instances was contacted for this information by the researcher. In a few instances some participants were not helpful citing reasons such as they did not know anyone who could be of value to the study. However, some interviewees introduced the researcher to more than one other person who was then contacted. It has been argued by Biernacki and Waldorf (1981) that in order to maximise sample variability and the theoretical utility of snowballing sampling, researchers need to make conscious efforts to pace and monitor the referral chains that they generate. Their viewpoint of snowball sampling is best seen as proceeding through a number of phases. Hence as part of the selection process, every referred person was communicated to in order to gauge their level of Internet usage for searching and purchasing products, before proceeding with organising an appointment for the interview. In at least three cases it was found that the referred person did not meet these criteria and hence was not included for an interview, the person was thanked for their time and co-operation.

Snowballing sampling technique is appropriate when research is concerned with a small, specialised population of people who are knowledgeable about the topics (Aaker and Day 1990; Patton 2002). A key issue in the snowball sampling technique is to be careful in the selection process of the first interviewee as the first snowball. In this study extreme care was taken to recruit the first interviewee who was one of the heaviest surfer and shopper amongst the twenty participants. Patton (1987; 2002) cites seminal work conducted by various business “gurus” that utilised the snowball sampling technique. Accordingly the best-known ones were The Peters and Waterman (1982) study In Search of Excellence, which used a snowball sampling technique, where the researchers asked a broad group of knowledgeable people to identify well-run companies. Rosabeth Moss Kanter’s (1983) study of innovation reported in The Change Masters.
focused on 10 core case studies of the “most innovative” companies. She began by asking corporate experts for candidate companies to study. These nominations then snowballed as she broadened her inquiry and then converged into a small number of different expert informants. Semi-structured in depth interviews were used to gather data relating to how people behave while using the Internet and websites, during the search and shopping activities online. The in-depth interviews were conducted over a period of three months, and lasted between one to two hours. They were audio taped, and the tape recordings were transcribed verbatim by the author, in doing so the contents of the interviews were reviewed several times.

Direct tape recordings of field events must be either transcribed fully, if the aim is to have a full record of the speech and other audible events, or processed in some way like the field worker listening to the recordings makes notes, selects excerpts, makes judgements or ratings and so on (Miles and Huberman, 1984). In this way, not only was the memory refreshed, but in-depth contextual understanding of the issues was also achieved. It was then synthesized into a word-processed database that was cross-referenced to the context of the interview. The transcription of these interviews took between four to six hours.

6. Discussion of the results – Motivators for consumers to search and shop for products online

The main motivating factors that drew respondents to the Internet to search and shop for products and services are as follows:
- Convenience
- Saving time
- Cheaper prices and also the ability to compare prices
- Good place to shop for specialty and hard to access products
- Level of control
- Ease of comparing products and services

6.1 Motivator - Convenience

The ‘convenience’ factor was cited as the main reason by all twenty interviewees, for using the Internet to search and shop for products. Some of their comments are as follows: Interviewee comments on motivator convenience:

“Convenience is the prime reason. Get it, Got it, Good I just get it and I am out of there. Convenience, more than anything, also, simply because I some time buy specialty type items and the Internet is convenient place for me to do my searching”.

“Very much convenience I should say so that I can shop when I want to shop which is very much after hours that is after normal trading hours so that is definitely the biggest motivator and feature for me”.

Though convenience was the top most motivating factor for people shopping online, the delivery timings adopted by Australia’s two major supermarkets i.e. Coles and Safeway were considered as major barriers to shop for grocery online. Most respondents mentioned that these supermarkets home delivered their produce between 9am – 5pm, Monday to Friday, which was not convenient for families especially if both partners worked.

6.2 Motivator - Saving time

The second most important reason for people to shop online was because it saved them time. Closely linked to convenience is the notion that the Internet is a “mouse click” away, and can actually save consumers a lot of time in their shopping or physical browsing. All of the respondents interviewed had fulltime office jobs and two of them also studied during the evenings after work. All were of the opinion that life generally was getting a bit hectic with work, study, and family responsibilities. The Internet has now become an extremely convenient gadget and made life less hectic. It is a time saving piece of equipment when it comes to most of their personal commercial chores. Interviewee comments on motivator time-saver:

“I find it a lot easier (Internet) instead of just browsing and doing window-shopping, which is wasting a lot of my time and then I could be doing something else out there. I would rather spend that extra dollar or two to get online find exactly what I want.

“Well with Coles online I did shop with them because I was running out of time and it was schools, studies, and tending to family”.

6.3 Motivator - Cheaper prices and also the ability to compare prices

The people interviewed had mixed feelings when it came to getting cheaper prices over the Internet. Some found that they were motivated because of cheaper prices, but at least half of the respondents felt that cheaper prices was not a motivating factor and in fact the Internet did not provide cheaper prices for products and services. However most agreed that the Internet was a useful place to make comparisons between prices of products and this was one of the reasons that
they searched and shopped online. Interviewee comments on motivator cheaper prices and comparison of prices:

“Definitely price. The whole idea over here is I can get better deals in airline travel”.

“With hotels and what I do because they are quite a few websites now that offer similar services. I would actually look at say three or four and have a look at their prices compare them. I have noticed that they (hotels) tend to be cheaper online”.

6.4 Motivator - Good place to shop for specialty and hard to access products

Another key motivating factor cited by interviewees, was to get information on specialty type products and services. This included products that are unavailable in Australia, but available in overseas markets. With the global reach of the Internet, consumers are now able to source these products via the Internet. Interviewee comments on specialty and hard to get type of products and services:

“If it isn’t readily available in Australia or outrageously expensive I will buy it online from overseas”.

“I am motivated to go online to buy specialty items so things like China, I bought this dinner set and I am looking always online for specific pieces. I will search interstate and overseas for a particular piece and that type of thing”.

“I recently bought a CD for my sister and it was an independent American artist, being such a unique and special artist the Internet was the only place I was able to get it being in Australia”.

6.5 Motivator - Level of control

The ease with which shoppers could obtain information was cited as an important reason for them to go online. This coupled with a feeling of consumers being more in control of information for purchase decision-making was given as key motivating factors to be online. The emotions with which consumers spoke about being in control of information to enhance purchase decision-making, is a point that should be noted with care and interest by marketers in designing communications through the company website.

Interviewee comments on motivator consumer control:

“Yes with more information at my finger tips I feel more in control. I find that I do better deals because there have been deals offered online which travel and tour operators cannot refuse and deny. As an example I recently drove a very good bargain with a Sydney hotelier with information I accessed online”.

“Probably the motivating factor is the ability to get fairly wide range of prices and information quickly. The speed and the breadth of information available online is incredible. Imagine trying to get the same amount of information from a Myer or a David Jones where it could take you a half or a full day to get the same amount of information”.

6.6 Motivator - Ease of comparing products and services

Consumers investigate and compare information between competing and substitute type of products and services online. This was also a key motivating factor for consumers to be drawn to the Internet. Interviewee comments on comparison products and services

“I found in the USA when I compared products online, I was able to get more features when I bought my computer. I was able to get a zip drive as part of the package from another wholesaler”.

“Basically I just browse around just to look and see what are the products that are available online and how businesses like Kmart, Myer, and Coles go about their sites.

7. Conclusion

This study found that ‘convenience’ was a key motivating factor for consumers in Australia to search and shop for products and services over the Internet. It was also found that searching and shopping for products over the Internet, saved consumers time, some suggested it saved them from the task of mundane shopping. Cheaper prices and a greater variety of products and services which were available from around the world online was another key motivating factor. Another key motivating factor included the ability for consumers to search and shop for hard-to-find products, most of these unavailable in Australia.
References

Sociological Methods and Research, 10, pp.141-163.
USA.
Thousand Oaks, California, USA.
European Business Review, 15/3, pp.141-152.
Riley Dall’Omo Francesca and Lacroix Caroline (2003) “Luxury branding on the Internet: lost opportunity or impossibility?” Marketing Intelligence and Planning, 21/2, 96-104.


A Case Study on the Selection and Evaluation of Software for an Internet Organisation

Pieter van Staaden¹ and Sam Lubbe²
¹Faculty of Informatics, Cape Peninsula University of Technology, Cape Town, South Africa
²School of Information Systems and Technology, University of KwaZulu-Natal, Durban, South Africa
²slubbe@ukzn.ac.za

Abstract: The authors conducted research to determine whether IT-managers, IT auditors, users, management, etc. (all decision-makers) use a certain evaluation and selection process to acquire software to meet business objectives and the requirement of users. An argument was used that the more thorough the software evaluation and selection process, the more likely it would be that the organisation will choose software that meets these targets. The main objective of the research was therefore to determine whether Media24 uses evaluation methods and obtain the desired results. Media24 is Africa's biggest publishing group and offers entertainment, information and education 24 hours a day and served as a good example of an Internet Service organisation that made their employees available for this research project. The results confirmed that Media24 uses suggested protocol as noted in the theory for software acquisition correctly during most stages.

Keywords: Black Box testing, business process, commercial software system, document, evaluation, request for proposal (RFP), requirements, selection, software, vendors.

1. Introduction

There is a wide variety of methods that can be used for selection of software in various fields of business (e.g. Manufacturing, Service providers, Insurance, Wholesale, Retail, etc.). This software is used for a variety of purposes in businesses. However, selecting the software that meets organisational requirements and business objectives could prove to be a challenge considering the number of vendors and software available.

“Choosing the right software for your company can be bewildering. There are thousands of titles to choose from, and programs and their functionality differ frequently” (Buyerzone.com, 2002).

A hurried, uneducated choice could lead to various problems in the company. Some of these are: failing to support an important business process; supporting a process inaccurately or inefficiently; unhappy customers; disgruntled employees; loss of sales; and poor financial performance. Competition in the Western Cape requires good performance in all aspects of the electronic publication industry. Bad judgments or decisions in terms of software acquisition could cause a company some losses and complications in a company’s daily operations. Choosing the right software is therefore important and can be achieved by using predetermined evaluation and selection guidelines.

2. Evaluation and selection of a commercial software system

2.1 Decisions made prior to the software evaluation process

As mentioned by Capterra’s Software Selection Methodology (2002), certain procedures should be completed before the actual evaluation is conducted. It suggests that the company should start off by interviewing some staff members, addressing corporate vision, analysing existing systems limitations and features and looking at present policies and procedures. The company should also determine whether new software will help the business and if it will increase competitive advantage.

It argues that when the decision is made to purchase software, a project plan should be developed to evaluate and list the evaluation criteria that will be used during the process. A project team should also be selected to carry out the evaluation. This team must include representatives from all levels the organisation. If the proposed software incorporates financial aspects the audit team should also be included.

2.2 Determine requirements for the new software Package

The purpose would be to create a comprehensive and prioritised list of requirements to help evaluate the software. Base Consulting Group (BCG) (2000) state that the requirements definition should consist of several processes
(such as: Managerial Requirements (Budget/Timing, Reporting Requirements), Functional Requirements (Stated Business Needs, Technical Requirements) and IS Standards (Data Flow Diagrams, System Interfaces, and Hardware and Network requirements with emphasis on capacity). They also note that some companies do not develop detailed requirements and as a result may be dissatisfied with the final outcome. Romney and Steinbart (2000) support this statement and suggest that one or any combination of four strategies (listed below) should be used to determine requirements for the new software:

- Surveying end-users to determine what their requirements for software is by using questionnaires, personal interviews and focus group discussions.
- Analysing the existing system and eliminate requirements that already has been defined.
- Examine how the existing software is used and helps to determine the shortcomings of the system and identify any new requirements needed by users.
- Pilot demonstrations of applications/software systems could be utilised when there is a problem in identifying requirements.

2.3 Document the requirements

Table 1: Capterra’s methodology (2002)

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional department, business</td>
<td>This creates the requirements hierarchy and ensures that all processes are</td>
</tr>
<tr>
<td>processes, and processes.</td>
<td>covered (e.g. would be creditors department, cheque printing, approving</td>
</tr>
<tr>
<td></td>
<td>cheques for printing, etc.)</td>
</tr>
<tr>
<td>Requirement type</td>
<td>Identifies the requirement as functional, technical, vendor related, or</td>
</tr>
<tr>
<td></td>
<td>contractual.</td>
</tr>
<tr>
<td>Requirement description.</td>
<td>This details the requirement itself and should be as descriptive as possible.</td>
</tr>
<tr>
<td>Priority and ranking</td>
<td>This could be used during the evaluation</td>
</tr>
<tr>
<td>Objective addressed</td>
<td>This could be used to match the requirement to a business objective.</td>
</tr>
<tr>
<td>Comments</td>
<td>This can be used for any additional comments or justifications</td>
</tr>
</tbody>
</table>

The systems requirement document or software requirement specifications should be the starting point for measuring performance, of the final system (Shelly et al., 1998). Users must understand the document to be able to improve the final version. The content of this requirements document will depend on the type of company and the complexity of the new system. BCG (2000) states that the requirements document is the cornerstone to evaluate the software and should be used to identify requirements. Capterra (2002) argues that there is a methodological approach available to help with requirement analysis. This is listed in Table 1 Capterra’s Methodology (2002).

3. Selecting vendors

3.1 Select list of possible vendors

Michell and Fitzgerald (1994) argue that the range of services offered by IT vendors is large and growing rapidly. They also note that while searching for the “best” vendor, it should be borne in mind that the process of selection and evaluation a vendor is important. BCG (2002) suggests that the project team’s first step should be to identify the vendors who offer software solutions that could be used. It could be a high-risk approach not to properly evaluate vendor companies. The sources used to compile the list of vendors should be recent, and reliable. These sources include software user groups, databases, industry groups, research firms, consulting firms, trade shows, seminars and conferences, current users, personal recommendations and contacts, competitors, IT and business magazines as well as Web sites.

4. Preliminary evaluation of vendor companies and their product

Ward (2001) argues that selecting software from a vendor should be a simplified process. She suggests that vendors be invited to participate in a software demonstration because it reduces time spent on evaluations. Companies could work toward solving the business problems earlier resulting in faster return on investment. BCG (2002) states that inviting too many vendors to participate increases the costs and timelines of the project. Team members may also lose focus after seeing too many product demonstrations. In order to shorten the time of the process send the Request for Proposal (RFP) to a shortlist of 5 vendors, and do a preliminary evaluation of the vendors. The Evaluation Team should look at things like:

- The standard functionality and key features of the product;
- Technology requirements (hardware, additional software, database, operating system, network, development tools);
- Product considerations such as viability, stability and cost;
- Products targeting different, smaller or much larger companies or industries should be eliminated;
5.2.2 Evaluation techniques and methods

Romney and Steinbart (2000) suggest benchmarking while processing times of software are measured. Software with the lowest time is normally judged the most efficient. Oberndorf et al. (1997) engaged scenario-based testing methods to represent typical procedures for the software to be programmed and not the software undergoing tests. Test procedures are developed based on scenarios and each is evaluated against a set of criteria. In this case the initial scenarios are established using preliminary operational definitions. The results of this will serve as confirmation that the software performed satisfactory against set parameters.

Romney and Steinbart (2000) suggest a point scoring technique to evaluate the vendor. Each criterion is assigned a weight based on its relevancy. The vendor is assigned a score based on how their proposal measures up to each criterion. The vendor with the highest score is then judged the 'best'. They argue that ‘requirements’ costing is an alternative where the total cost of the proposed software is calculated. This provides an equitable basis for comparison. Another method suggested by Voas et al. (1997) is the use of fault injection techniques. This is effective when buyers do not have access to the source code. The method consists of inserting erroneous values into the control stream and checking the results. This technique is an example of evaluating (for discovery) to determine unknown or unexpected reactions of the product under evaluation.

Beizer (1995) suggests Black Box testing to allow a tester to treat each module as a unit which can be defined by its inputs and outputs (the interfaces to the module) without considering the route by which an input is transformed into a particular output. Visibility into the internal workings of the code module is not necessary and source code not required. An example of black box testing is boundary value analysis where inputs are supplied to the software to be tested (these values represent valid, invalid and parameters). The outputs are measured and accepted if they fall within expected limitations. This type of testing is used during acceptance testing and is the basis of validation testing, confirming the software performed the required functions.

Other techniques (Hausen and Welzel, 1993) include analysing product documentation, presentations, using trial versions, scheduling demonstrations or attending training of the software. They suggest that one or, more of the above-mentioned techniques could be used to
supplement the evaluation of software. The project team should use discretion when selecting evaluation techniques as a company's approach and resources may vary.

5.2.3 Evaluation considerations

Hausen and Welzel (1993) mention that some of the following principles should be taken into consideration:

- Repeated testing of the same product using the same product specifications with the same testing techniques must deliver similar results (Repeatability).
- Repeated evaluation of the same product to the same product specifications by different parties must deliver similar results (Reproducibility).
- The evaluation is free from bias while achieving any particular result (Impartiality).
- The result is obtained with minimum subjective judgment (Objectivity).

5.2.4 Product evaluation

Hausen and Welzel (1993) state that the evaluation process should consider software features (compared to the requirements document), product information (acquired from the RFP, product demonstration, information gathered from investigating vendors, etc.), evaluation techniques and process information (e.g. results obtained from the testing techniques).

Capterra's methodology (2002a) states that all software should be evaluated to determine if it meets requirements (functional and technical). Any additional (functional/technical) requirements should be listed and re-calculated. Missing requirements should be listed and cost incurred to add these features should be calculated. Price and maintenance levels of the product have to be evaluated by totalling cost and maintenance levels. Firms must consider initial product costs (also long-term costs (such as training, implementation costs, maintenance, and upgrading costs)). Project teams should keep in mind that software is expensive and by picking the wrong one could have costly repercussions.

5.2.5 Final evaluation of vendor companies providing possible software solution

Pollard (1999) suggested the evaluation of the support and maintenance staff of the vendor. He notes that it is necessary to know the number of people in customer support. The response time can be measured by calling the customer support department. The availability and quality of the implementation support also ought to be evaluated. The new software could have bugs and other problems (e.g. not meeting the required deadline). The vendor must provide training because users want to use the system properly.

He suggests following up on customer references, reviewing case studies and finding out how many companies are using the software. The financial stability of a vendor is an aspect to consider. Pollard (1999) supports this by suggesting the examination of the financial history and the long-term financial stability of the vendor.

5.2.6 Selecting the software system

The total score of the software should be recorded on a scoring sheet when a point scoring technique is used. All the software must be listed from the highest to the lowest. The software with the highest score would represent the best fit for the organisation. Although the software with the highest score might represent the best solution for the company, there may be reasons unrelated to the requirements that could prevent an organisation from selecting software. Inconsistencies should be identified (also extremes in scoring that may influence it - or a competency or deficiency within a single business function). Criteria such as a business partnership, potential future business, or other intangibles, must also be included.

5.2.7 Notify the vendor

Once the steering committee has approved the vendor, then the vendor should be notified and a contract drawn up. The diagram (on the next page) was derived from the theory to illustrate steps used during the evaluation and selection process.

6. Problem statement

Brown and Wallnau (1996) state that organizations should recognize the importance of technology 'refreshment':

- To improve the quality of their products and services;
- To be competitive with organizations providing similar products and services; and
- To remain attractive to investors.

Any organization should invest in appropriate software to stay in business. Careful decision-making on investment into software is therefore essential. Whether the release of an update, or the availability of new software, should force an organization to initiate an evaluation process that provides timely, balanced information on which decisions can be made. The problem statement was thus stated as:
...the more thorough the evaluation of software, the greater the chances could be for the organization to select the software that will meet their business objective and the requirements of the users.

All the methodologies listed above were summarised in the figure in Appendix 1 at the end of the paper.

7. Research methodology

7.1 Objective

The objective of the study is to determine whether Media24 uses the correct software evaluation and selection guidelines when purchasing software from a vendor as prescribed in the theory and whether these guidelines have obtained results. Media24 is Africa's biggest publishing group and offers entertainment, information and education 24 hours a day (www.media24.com; www.news24.com, etc.). Based on the theory and the problem statement, the research questions were stated as follows:

- Is the software evaluation process of Media24 thorough enough to select software that fulfils their end-user requirements?
- Does the organisation pick software that meets the business objectives easily?

7.2 Limitations of the study

The researchers had decided to exclude legal procedures and only focussed on the project from a technical viewpoint. Also they were not allowed to use the name of any software used by Media24 as it might discredit the vendor or Media24. The research scope also does not include the processes (e.g. contract negotiation) followed after a decision has been made.

7.3 Development of questionnaire

The researchers compiled a questionnaire consisting of two sections (Section A and B). Section A contained 7 sub-sections, covering aspects of evaluation and selection processes. Section B contains the user satisfaction survey. The questionnaire was tested for any questions that might cause a problem with 5 academics from Cape Technikon who was not included in the sample. Any vague questions were also improved or removed. The questionnaire was handed to a Media24 IT Manager who reviewed it. He also identified people responsible for evaluating and selecting software in Media24 to be respondents while he completed a questionnaire himself.

Section A was sent out to 15 decision makers involved in acquiring software in Media24. The objective was to measure whether they used the evaluation techniques as prescribed in the theory. Section B was sent to 50 users of the software. The objective was to determine whether they were satisfied with present software. All questions where derived from the theory described earlier. A five point Lickert scale was used in most of the questions (it included questions that required respondents to pick more than one answer). The reason is to evaluate areas where answers from the respondent indicates an in depth approach to the evaluation process. At the end of each section a section was dedicated to find out in more detail how respondents think.

7.4 Analysis of data

The results of both sections were compared using cross tabulation. The data was captured and analyzed on an Excel spreadsheet. Twelve ‘Section A’ respondents returned the questionnaires and 74% of the respondents who received section B returned the questionnaire. It was decided not to conduct any association analysis as the mode of the study did not require it at this stage. It was rather decided to conduct a follow-up study at a later stage when the student will be employed at Media24.

7.4.1 Section A: Evaluation process

<table>
<thead>
<tr>
<th>Table 2: Project management structure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>There is a Steering Committee in place to oversee and direct the evaluation process</td>
</tr>
<tr>
<td>Sufficient resources are allocated to the evaluation process</td>
</tr>
<tr>
<td>A project team with the required expertise is assigned to conduct the evaluation and selection of the new software system</td>
</tr>
<tr>
<td>Average</td>
</tr>
</tbody>
</table>

Nearly all the respondents agree that the Steering Committee correctly directs the evaluation process (11). This indicates that people realise that there is a structure in place that could oversee the evaluation of software and is in agreement with a similar comment by Capterra (2002). The process in Media24 is listed below (Table 2):

All the respondents agree that there are sufficient resources available to ensure that the evaluation
process runs smoothly. Nine of the respondents indicated that the project team has enough expertise to conduct an evaluation. Three have indicated that they disagree with this. There might be a problem that falls outside the scope of this study and needs to be addressed by management. The organisation therefore needs to assemble a project team that is representative of the people working in the organisation (see also Dean and Vigder, 2002). Nine of the respondents indicated that Media24 used the interview method to determine requirements. The same respondents have also used the present system to ensure that they meet the correct requirements (see Figure 4 later). Six respondents have indicated that they use questionnaires to collect requirements. It seems that project teams in Media24 prefer to use three methods to determine software requirements.

All respondents have indicated that technical requirements are determined beforehand (systems interface). Ten of the respondents are happy that business objectives have been met while determining functional requirements. All agree that managerial requirements should be met when requirements are defined. Nine of the respondents noted that requirements are properly documented. Media24 needs to address this to ensure that people are in agreement otherwise it can become a problem. The requirements definition as applied in Media24 (as reflected by the respondents) is listed in Table 3 below.

Table 3: Requirements definition

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neither agree or Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical requirements are determined (e.g. the system interface)</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>The Functional requirements are determined (e.g. business objectives system has to fulfill)</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Managerial requirements are determined (e.g. reporting capabilities, budget, timing)</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Requirements are properly documented and easy to understand</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Average</td>
<td>4</td>
<td>7</td>
<td>.5</td>
<td>.5</td>
</tr>
</tbody>
</table>

Ten of the respondents agree that many sources should be investigated to identify correct software. The same number agrees that Media24 should ensure that software that meets requirements is evaluated. This is done to limit the number of products to be considered. On the other hand, 11 of the respondents agree that the support/maintenance staff provided by the vendor is evaluated on response time while the quality of support is rated high by 10 of the respondents. It seems to the authors that ‘old fashioned’ values are still important while looking at new IT investments. This is in agreement with the statement made by Capterra (2000) that end-users value help provided by the supplier. The importance of vendor identification by Media24 is reflected in Table 4.

Table 4: Vendor identification and evaluation

<table>
<thead>
<tr>
<th>Source of Information</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neither agree or Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Various sources (e.g. web sites) investigated in order to identify the software available.</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>A preliminary evaluation is conducted to limit software that is going to be extensively evaluated.</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>The support and maintenance staff provided by the vendor is evaluated based on:...</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...response time</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...quality of support</td>
<td>1</td>
<td>9</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>...number of people</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>...cost</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>The company providing the software system is evaluated based on:...</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>...long term financial stability</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>...customer references</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...number of clients</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>...long and short term strategic planning</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Average</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

The respondents do not regard the number of people working for the vendor as important; as long as their service is not affected by it (quality is...
rated higher). Eight stated that cost plays a role while maintenance is evaluated. Most of the respondents noted that the vendor should be evaluated on financial stability and customer references. The number of clients and long term strategy of the vendors were not used when looking how reputable the vendor is. This is something that Media24 and similar organisations need to investigate.

The importance of a request for proposal is listed in Table 5.

Table 5: Request for proposal

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neither Agree or Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Request for proposal is sent to Vendors.</td>
<td>10</td>
<td>1 1 0 0</td>
<td></td>
</tr>
<tr>
<td>The RFP includes…….</td>
<td>1</td>
<td>9 2 0 0</td>
<td></td>
</tr>
<tr>
<td>the purpose of the RFP</td>
<td>4</td>
<td>7 1 0 0</td>
<td></td>
</tr>
<tr>
<td>...all necessary background information of the company issuing the RFP</td>
<td>1</td>
<td>4 6 1 0</td>
<td></td>
</tr>
<tr>
<td>...the requirements for the new system</td>
<td>3</td>
<td>4 4 1 0</td>
<td></td>
</tr>
<tr>
<td>...the hardware and software environment currently being used</td>
<td>9</td>
<td>2 1 0 0</td>
<td></td>
</tr>
<tr>
<td>...instructions on how to reply and a description of how responses will be dealt with</td>
<td>1</td>
<td>9 2 0 0</td>
<td></td>
</tr>
<tr>
<td>...a statement of work (SOW) that includes the required work needed for the procurement of the new software solution.</td>
<td>1</td>
<td>9 2 0 0</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>4 5 2.6 .4 0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 displays the data collected on the RFP. Most of the respondents agree that this is complied with by Media24. Many (10) stated that the evaluation processes are properly documented while thought is given to objectivity (11) and impartiality (9). Most (10) of them agree that repeatability is lacking while reproducibility (7) is important and needs careful attention. Impartiality is an aspect that needs to be addressed as a third of the people noted that it is lacking. The averages for the request for proposal can be improved but because there are individual items that management of Media24 needs to pay attention to as stated. The common method used to evaluate the software is the point scoring technique (Figure 2). Trial versions of software have a better chance to be selected if it was analysed previously or used. The other method that Media24 uses to evaluate software is the expertise of their staff in a particular field.

Figure 2: Techniques used while evaluating software

Table 6 shows the evaluation process. The respondents were positive that the final score used to determine how well the software product meets the requirements is a good method. Nine respondents are in agreement that the documentation to do this evaluation is well laid out beforehand. Management needs to convince the 3 that does not agree to accept the documentation as presented. Objectivity is complied with but repeatability of the evaluation is not highly thought off (and the same for reproducibility). Nine of the respondents noted that the tests were impartial. Again it would be a task for management to convince the remaining three that this is the case before problems are experienced (also supported by Shelley et al. (1998)).

Table 6: Evaluation of the product

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neither Agree or Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The evaluation score of the product is determined based on how well it meets the predetermined requirements.</td>
<td>1</td>
<td>11 0 0</td>
<td></td>
</tr>
<tr>
<td>The evaluation process and results are properly documented</td>
<td>1</td>
<td>8 2 1 0</td>
<td></td>
</tr>
<tr>
<td>When conducting the evaluation thought is given to……….</td>
<td>7</td>
<td>4 1 0 0</td>
<td></td>
</tr>
<tr>
<td>...Objectivity</td>
<td>0</td>
<td>0 4 7 1</td>
<td></td>
</tr>
<tr>
<td>...Repeatability</td>
<td>0</td>
<td>6 4 2 0</td>
<td></td>
</tr>
<tr>
<td>...Reproducibility</td>
<td>7</td>
<td>2 2 1 0</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>3 5 2 1.8 .2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Most of the people agree that the final decision to acquire software is based upon the evaluation results (Table 7). Eleven of the respondents agree that they are happy with the results achieved by Media24. Agreeing meant that they use the
methods as prescribed in theory but there are some methods that are not used presently (e.g. benchmarking and black box testing). These should be investigated and used to ensure that the methods presently being used are still considered the best (Beizer, 1995).

Table 7: Results obtained from decision makers

### 7.5 User satisfaction and overall effectiveness of evaluation process used

Nearly all the respondents (11) at Media24 are satisfied with the software that was purchased by Media24. Eight of the respondents have agreed that it falls within the parameters set by the organisation (as displayed in Table 8). The rest that does not agree (4) may be users that were not part of the project team. They should be convinced that the software is of benefit to Media24. Eleven of the respondents agreed that the software meet their requirements. This is supported in their article by Michel and Fitzgerald (1997) who stated that normally most of the users who were part of the process are happy with the software.

Table 8: User satisfaction

<table>
<thead>
<tr>
<th></th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neither agree or Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am satisfied with the software I am using.</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>The commercial software system fulfills the business objective it was assigned to</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>The commercial software system adheres to my requirements</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>There is no room for improvement for the commercial software system I am using</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Average</td>
<td>1.5</td>
<td>7</td>
<td>1.25</td>
<td>2</td>
<td>.25</td>
<td>12</td>
</tr>
</tbody>
</table>

This could mean that the individual user requirements agree with Media24’s requirements. However, 7 of the respondents agree that there is room for improvement. This should be investigated in another study as this falls outside the scope of this study. This does not agree with previous statements made by the respondents. Maybe the present evaluation procedure should be extended to inquire about reasons why the respondents argue that there is room for improvement. This should be included as part of the evaluation and is displayed in Figure 3.

**Figure 3:** Number of complaints from users

The respondents (9) complain about the software occasionally to ensure that management pay attention. Seven of the users argue that there is room for improvement. This would require another survey to find out why users complain about the software and if there is room for improvement.

### 7.5.1 Section B

The main objective of the section B was to assess whether the users where satisfied with the software acquired by Media24. There are some complaints lodged by the respondents (e.g. the systems response time and redundant processes and procedures included in the current system and that the software does not integrate with other software). The figure on the next page illustrates the percentage respondents who agreed that they use the correct evaluation techniques. The results are summarized according to key areas in Figure 4.
The researchers note that there were some complaints mentioned by end-users. They feel that this ‘unhappiness’ could be because not all employees were actively involved in determining the requirements or because the requirements were not communicated to all users. A manager, during an interview, noted that the software he was using does not integrate with one of the sub-systems. This eventually leads to more work, as manual reconciliation has to be done between the sub-system and the software. This could be because the requirements or the software weren’t described properly in the Request for Proposal. Most of the statements indicated that e-Commerce organisations should be careful how they select software (Capterra (2002), Hausen and Welzel (1993) and Lars and Methven (2002)). This was also supported by the research findings of this study.

9. Future research

Further research might be needed to refine or redesign the evaluation approach used by Media24. The reason is that the current software will be outdated soon and with the rise of new technology the evaluation and selection process might have to be adjusted. There were also some issues that were not picked up before the installation of the product (e.g. integration with sub systems). This indicates why the evaluation and selection process might have to be revised. Other studies could help place emphasis on the use of specific evaluation models. In order to speed up the process and to gather more data more respondents will have to be included in the sample before the next survey is conducted.

References


Appendix 1: Graphical Summary of Software Selection