Developing Information Systems Design Knowledge: A Critical Realist Perspective

Sven A. Carlsson
Informatics, School of Economics and Management, Lund University, Sweden
sven.carlsson@ics.lu.se

Abstract: Academic Information Systems (IS) research has a serious utilization and relevance problem. To increase IS research utilization and relevance, scholars argue that the mainstream IS research, which is based on the behavioral science paradigm, should be complemented with research based on the design science paradigm. The current IS design science frameworks have a strong focus on the IT artefact, in most cases an exclusive focus on the IT artefact. The frameworks have very little discussion and clarifications regarding underpinning philosophies, but most seem to be based on positivism, traditional realism, or pragmatism. This paper presents an alternative framework for IS design science research. The framework builds on that the aim of IS design science research is to develop practical knowledge for the design and realization of different classes of IS initiatives, where IS are viewed as socio-technical systems and not just IT artefacts. The underpinning philosophy of the framework is critical realism which has been developed as an alternative to positivism and traditional realism as well as to constructivism (relativism). The framework proposes that the output of IS design science research is practical IS design knowledge in the form of field-tested and grounded technological rules. The IS design knowledge is developed through an IS design science research cycle. The paper presents how technological rules can be developed as well as the nature of such rules.

Keywords: Information systems, IS design, frameworks, rules

1. Introduction

In the last years we have seen an intensive debate in the Information Systems (IS) community on the “crisis in the IS field”—see, for example, the debates in the journals MIS Quarterly and Communications of the AIS. Some commentators argue that part of the crisis is related to the utilization and relevance problem (Hirschheim and Klein, 2003): research not addressing relevant issues and research not producing usable results. To increase IS research utilization and relevance it is argued that the mainstream IS research, which is based on the behavioral science paradigm, should be complemented with research based on the design science paradigm (Walls et al., 1992, 2004; March and Smith, 1995; Hevner et al. 2004).

Research can be divided and classified in different ways. Herbert Simon (1988) in his seminal book “The Sciences of the Artificial” distinguish between “natural sciences” and “sciences of the artificial”. The former focuses on how “things” (natural and social things) are and how they work—for clarity and consistency we will in the rest of the paper use the concept “behavioral science” instead of “natural science”. The sciences of the artificial focus on how to make artefacts and artificial systems having desired properties. Even if it is common to think of engineers, architects, and industrial designers as typical professional designers, Simon stresses that “Everyone designs who devises courses of action aimed at changing existing situations into preferred ones.” Simon’s work on the sciences of the artificial and design science has influenced IS scholars and we have in the last years seen a growing interest in IS design science research and IS design theory/knowledge (Walls et al., 2004); and there is also a fairly new ISWorld web-site on “Design Research in Information Systems” (Vaishnavi and Kuechler, 2005). Interesting IS design science research frameworks have emerged, but from our perspective two major issues have not been carefully addressed. First, there is too little discussion about what IS design science research should include and what should be excluded. This is related to the discussion about what the IS discipline ought to be and what ought to be at the core of the IS discipline. When there is a discussion the view hold is that it is IT artefact design theories that should be developed. Simon’s view on design science points to that it can be more than the IT artefact that the IS field should develop design knowledge for. We will argue that there is a need for IS design science research frameworks having a broader view on IS and IS design knowledge. Second, there is no, or little, discussion about underlying philosophical assumptions in the IS design science research literature. The underlying ontological view an IS design science research framework is built on will ultimately affect how to do IS design science research and what types of outcomes (design knowledge) that can be produced. Although, current frameworks lack in clearness on underpinning philosophies and ontological views, they seem to be based on positivism,
traditional realism, or pragmatism. In IS research based on the behavioral science paradigm there is an increased and fruitful use of alternative philosophies (ontologies), for example, the use of constructivism and critical theory. Consequently, we suggest that it can be fruitful to develop and explore IS design science research frameworks based on alternative philosophies (ontologies).

The remainder of the paper is organized as follows: the next section reviews IS design science research frameworks and elaborates the above two issues. The section argues for a broader view on IS design science research and for grounding IS design science research in the philosophy of critical realism. A short presentation of critical realism follows and this is followed by a presentation of an IS design science research framework based on the philosophy of critical realism. Our work builds on the idea of Pettigrew’s (1997) primary double hurdle: IS design science research should meet the criteria of scholarly quality and practical (professional) relevance.

2. A review of Information Systems design science research frameworks

Simon’s distinction has influenced the IS field. For example, Järvinen (2004) distinguishes between research approaches stressing “what is reality” (behavioral science) and research approaches stressing “utility of artefacts” (design science). Although using different concepts, Walls et al. (1992), March and Smith (1994), and Hevner et al. (2004) make similar distinctions.

Behavioural science is description- and explanatory-driven whereas design science is prescription-driven. Simon argues that there has been a movement towards behavioral science and away from the design sciences in engineering, business and medicine. Although, the IS field is quite young and Simon’s book was first published in 1969, reviewing articles published in the leading IS journals reveals a picture where the majority of published articles belong to the behavioral science paradigm (Glass et al., 2004; Chen and Hirschheim, 2004). Based on the IS field’s utilization and relevance problem it has been suggested that one way to advance the IS field is to increase IS design science research (Hirschheim and Klein, 2003; Liivari, 2002).

Below we review IS design research frameworks by primarily focusing two issues:

1) what are included in the IS design science research frameworks, and 2) what underlying philosophies (ontologies) have the frameworks. The first issue is related to the discussion on what the IS discipline ought to be and what ought to be at the core of the discipline. The second issue is critical since in all research, including IS design science research, ontology is non-optional (Fleetwood, 2004).

As far as we know, the first article on developing IS design theories (ISDT) and IS design knowledge was published in 1992 (Walls et al., 1992). Walls et al. argue that successful construction of ISDT would create an endogenous base for theory in the IS discipline, and could be used by scholars to prescribe design products and processes for different classes of IS as they emerged. The authors build on Simon’s distinction and argue that design is both a product and a process, which means that a design theory must have two aspects—one that deals with the product of design and one that deals with the process of design. Using their framework the authors proposed an ISDT for the IS-class “Vigilant Information Systems.” The components of an IS design theory are summarized in Table 1.

Walls et al. use the concept “artefact” quite freely, but in reflecting on their 1992-paper they say: “We did not use the current phrase “IT artefact”, but in essence it was that to which we were referring.” (Walls et al., 2004).

Building on Simon’s ideas, March and Smith (1995) distinguish between design sciences and natural sciences. The former involves building and evaluating: 1) constructs which are “concepts with which to ... characterize phenomenon”, 2) models that “describe tasks, situations, or artefacts”, 3) methods as “ways of performing goal directed activities”, and 4) instantiations which are “physical implementations intended to perform certain tasks”.

Table 1: Components of an IS design theory (Walls et al., 1992)

<table>
<thead>
<tr>
<th>Design Product</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Meta-requirements</td>
<td>Describes the class of goals to which the theory applies</td>
</tr>
<tr>
<td>2. Meta-design</td>
<td>Describes a class of artefacts hypothesized to meet the meta-requirements</td>
</tr>
<tr>
<td>3. Kernel theories</td>
<td>Theories from natural or social sciences governing design requirements</td>
</tr>
</tbody>
</table>
Hevner et al. (2004), building on March and Smith, present a design science framework and guidelines around building and evaluating IT artefacts—Figure 1 depicts their IS research framework.

Hevner et al. expressed their view on what constitute good IS design science in the form of seven guidelines. The authors contend that each of the guidelines should be addressed in some manner for IS design science research to be complete. Guideline one—“design as an artefact”—says: “Design-science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation.” (Hevner et al. 2004, italics added to indicate similarity with March and Smith’s output view). And, the “result of design-science research in IS is, by definition, a purposeful IT artefact created to address an important organizational problem. … Our [Hevner et al.’s] definition of IT artefacts is both broader and narrower [than other IT artefact definitions] … It is broader in the sense that we include not only instantiations in our definition of the IT artefact but also the constructs, models, and methods applied in the development and use of information systems. However, it is narrower in the sense that we do not include people or elements of organizations in our definition nor do we explicitly include the process by which such artefacts evolve over time.”

Regarding what should be included in an IS design research framework, and consequently in IS design theory and knowledge, it is clear that Walls et al., March and Smith, and Hevner et al. have the IT artefact as the subject. They exclude the non-technological context by excluding people and organizations. Given, the frameworks’ focus and what the exclude the framework might better be named IT design science research frameworks.

Benbasat and Zmud (2003) suggest that the core of the IS discipline and IS research should be the IT artefact—a narrow view on the IS discipline and IS research. Alter (2003) suggests a broader view and argues that the core of the IS discipline should be “work systems”. The above IS design science research papers have views more in line with Benbasat and Zmud’s view than with Alter’s view. In the IS core debate, Myers (2002) argues for that the IS discipline is nowhere near ready to define an IS core—he argues for open, flexible, and adaptive views. Hence, he argues for broad and emergent views on the IS core. Said Myers: “I believe that diversity is a positive attribute and ensures the continued viability of the field in a rapidly changing environment.” (Myers, 2002) We agree entirely with Myers. It should be noted that Walls et al. and Hevner et al. say that IS design theories and framework can encompass more than the IT artefact. Furthermore, Hevner et al’s second design guideline—problem relevance—states: “The objective of design-science research is to develop technology-based solutions to important and relevant business problems”. (Hevner et al. 2004). It can be noted that lists ranking current and future critical IS-issues, for example published by the Gartner Group, often have non-technological issues as the most critical (relevant) and less easy to solve issues, for example, “how to align our business strategy and IT strategy”.

Our view is that an IS design science research framework should be based in what should be produced, that is, what kind of knowledge should be developed. We suggest that the aim of IS design science research is to develop practical knowledge for the design and realization of “IS initiatives” or to be used in the improvement of the performance of
existing IS. By an IS initiative we mean the design and implementation of an intervention in a social-technical system and where IS (including IT artefact) is a critical mean for achieving the desired outcomes of the intervention.

The second issue we address is underpinning philosophies and ontologies of the frameworks. The above IS design science research papers do not explicitly address ontology, but ontology is non-optional in all research (Fleetwood, 2004). Although, the above papers do not address underpinning philosophies and ontologies, it is possible to conclude that they are based in positivism, traditional realism, or pragmatism. This conclusion is based on the few philosophical and philosophy of science references used by the authors and that they use words like “prove”; and Hevner et al. explicitly refers to pragmatism. The ISWorld web-site on “Design Research in Information Systems” has a section on the “philosophical grounding of design research” (Vaishnavi and Kuechler, 2005). Unfortunately, the authors mix concepts and definitions and their use of key concepts are inconsistent with what is found in the philosophy and philosophy of science literature. For example, they say that “ontological and epistemological viewpoints shift in design research as the project runs through circumscription cycles ... This iteration is similar to but more radical than the hermeneutic processes used in some interpretive research.” This means that in IS design science research a researcher’s assumptions about how the world is “constructed” should change during a research project! What the authors probably mean is that our knowledge of the world changes which is quite a different matter. They also make what Bhaskar (1978) terms an “epistemic fallacy” in that they transpose what is an ontological matter—concerning what exists—into an epistemological matter of how to develop reliable knowledge. It is interesting to note that the authors make a reference (using Mario Bunge’s work) to critical realism: “Bunge (1984) implies that design research is most effective when its practitioners shift between pragmatic and critical realist perspectives, guided by a pragmatic assessment of progress in the design cycle.” (Vaishnavi and Kuechler, 2005). Unfortunately, they do not explore Bunge’s view.

To summarize, papers on IS design theory and knowledge or IS design science research do almost never explicitly discuss ontological issues and underpinning philosophies, but most papers seem to be based in positivism, traditional realism, and pragmatism. This is consistent with studies on publications in the IS field. The overwhelming majority of papers are based on a positivistic philosophy (Chen and Hirschheim, 2004). IS research commentators point out weaknesses in positivism, etc., and suggest the use of alternative philosophies, like constructivism and critical theory—for examples, see the chapters in Minglers and Willcocks (2004) and Whitman and Wosczynski (2004). This paper articulates a view on IS design science research based on the philosophy of critical realism which is an alternative to positivism as well as to constructivism.

3. Critical Realism

Critical realism (CR) was developed as an alternative to positivism (empiricism) and as an alternative to non-positivism, e.g. constructivism (relativism). The most influential writer on critical realism is Roy Bhaskar (1978, 1989, 1998, 2002). Unfortunately, Bhaskar is an opaque writer, but clear summaries of critical realism are found in Archer et al. (2000) and Chapter One in Bhaskar (2002).

Critical realism can be seen as a specific form of realism. Its manifesto is to recognize the reality of the natural order and the events and discourses of the social world. It holds that “we will only be able to understand—and so change—the social world if we identify the structures at work that generate those events or discourses ... These structures are not spontaneously apparent in the observable pattern of events; they can only be identified through the practical and theoretical work of the social sciences.” (Bhaskar, 1989). Bhaskar (1978) outlines what he calls three domains: the real, the actual, and the empirical (Table 2). The real domain consists of underlying structures and mechanisms, and relations; events and behavior; and experiences. The generative mechanisms residing in the real domain exist independently of, but capable of producing, patterns of events. Relations generate behaviors in the social world. The domain of the actual consists of these events and behaviors. Hence, the actual domain is the domain in which observed events or observed patterns of events occur. The domain of the empirical consists of what we experience; hence, it is the domain of experienced events.
Table 2: Ontological assumptions of the critical realist view of science (Bhaskar, 1978)

<table>
<thead>
<tr>
<th></th>
<th>Domain of Real</th>
<th>Domain of Actual</th>
<th>Domain of Empirical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mechanisms</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Experiences</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Xs indicate the domain of reality in which mechanisms, events, and experiences, respectively reside, as well as the domains involved for such a residence to be possible.

Bhaskar argues that; “...real structures exist independently of and are often out of phase with the actual patterns of events. Indeed it is only because of the latter we need to perform experiments and only because of the former that we can make sense of our performances of them. Similarly it can be shown to be a condition of the intelligibility of perception that events occur independently of experiences. And experiences are often (epistemically speaking) ‘out of phase’ with events — e.g. when they are misidentified. It is partly because of this possibility that the scientist needs a scientific education or training. Thus I [Bhaskar] will argue that what I call the domains of the real, the actual and the empirical are distinct.” (Bhaskar 1978). Critical realism also argues that the real world is ontologically stratified and differentiated. The real world consists of a plurality of structures and generative mechanisms that generate the events that occur and do not occur. From an epistemological stance, concerning the nature of knowledge claim, the realist approach is non-positivistic which means that values and facts are intertwined and hard to disentangle.

The literature on the philosophy of science discusses the differences between positivism, constructivism, and critical realism; for example, discussions on their ontological views. Good discussions in terms of doing real world research based on the different philosophies of sciences are available in Robson (2002) and Bryman (2001). Table 3 summarizes a critical realism view of science.

Table 3: A realist view of science (Robson, 2002)

<table>
<thead>
<tr>
<th></th>
<th>rational criteria.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Explanation is concerned with how mechanisms produce events. The guiding metaphors are of structures and mechanisms in reality rather than phenomena and events.</td>
</tr>
<tr>
<td>4</td>
<td>A law is the characteristic pattern of activity or tendency of a mechanism. Laws are statements about things that are ‘really’ happening, the ongoing ways of acting of independently existing things, which may not be expressed on the level of events.</td>
</tr>
<tr>
<td>5</td>
<td>The real world is not only very complex but also stratified into different layers. Social reality incorporates individual, group and institutional, and societal levels.</td>
</tr>
<tr>
<td>6</td>
<td>The conception of causation is one in which entities act as a function of their basic structures.</td>
</tr>
<tr>
<td>7</td>
<td>Explanation is showing how some event has occurred in a particular case. Events are to be explained even when they cannot be predicted.</td>
</tr>
</tbody>
</table>

Critical realism is a well-developed philosophy of science, but on the methodological level it is less well-developed. The writings of Derek Layder (1993, 1998) and Mansor Kazi (2003), as well as some of the chapters in Ackroyd and Fleetwood (2000) and Fleetwood and Ackroyd (2004), can serve as guidelines for doing research based on critical realism. Unfortunately, from an IS design science research perspective, most of the writings have been in the behavioral science paradigm, i.e., for theory development and theory “testing”.

Critical realism has influenced a number of social science fields, e.g., management and organization studies. With few exceptions, CR is almost invisible in the IS field. Mingers (2004), Mutch (2002), Carlsson (2004), and Dobson (2001) argue for the use of critical realism in IS research and discuss how this can overcome problems associated with positivism and constructivism. The writings on CR in IS have been focusing on the use of CR in the behavioral science paradigm and not in the design science paradigm. This paper uses CR as an underpinning philosophy for IS design science research.

4. Developing Information Systems design knowledge

This section presents and discusses an IS design science research framework based on
Critical realism. It starts with discussing what types of knowledge should be produced and for whom. This followed by a presentation of how IS design knowledge can be produced.

4.1 For whom should IS design research produce knowledge?

The primary constituent community for the output of IS design research is the professionals in the IS field. This means primarily professionals who plan, manage, design, build, implement, operate, maintain and evaluate different types of IS. The developed IS design knowledge is to be applied by individuals who have received formal education (or a similar training) in the IS field. An IS-professional can be defined as a member of a fairly well-defined group who solves real-world IS-problems with the help of skills, creativity and (scientific) IS-design knowledge. (For simplicity we call the problems IS-problems although it is more correct to say that someone has defined a problem where one, for one reason or another, has decided to solve the problem with an IS-initiative). Another important community is IS education, which means that the knowledge should be useful in different types of IS study programs.

Although, the primary community works in organizations driven by profit (utility) “maximization”; it should be stressed that CR also has a critical and emancipatory component. The frameworks discussed in section 2 have a clear management perspective and certainly not an emancipatory or critical stance. The emancipatory and critical issue is important, but here we just note it and leave the issue for further exploration and development.

4.2 What types of IS design knowledge can IS design research produce?

IS design science research should develop practical knowledge to solve a class of IS-problems. This means the development of abstract knowledge that can be used in designing and implementing IS initiatives. It is abstract in the sense that it is not a recipe for developing the X-IS-initiative for the Z-organization. A user of the abstract design knowledge has to “transform” it to fit the specific situation and context. The knowledge takes the form of field-tested and grounded technological rules—will be discussed below.

Following Pelz (1978), we can distinguish between conceptual and instrumental use of science. The former involves using knowledge for general enlightenment on the subject in question and the latter involves acting on research results in specific and direct ways. Although both types are relevant we focus on the development of design knowledge for instrumental use.

Using van Aken’s (2004) classification we can distinguish three different designs an IS professional makes when developing an IS-initiative: 1) an object-design, which is the design of the IS intervention (initiative), 2) a realization-design, which is the plan for the implementation of the IS intervention (initiative), and 3) a process-design, which is the professional’s own plan for the problem solving cycle and includes the methods and techniques to be used to design the solution (IS intervention) to the problem. IS design science research should produce knowledge that can be used by the professionals in the three types of designs.

4.3 Design knowledge as field-tested and grounded technological rules

Following Bunge (1967), we can say that design science research aims at developing stable norms of successful human behavior, i.e. rules. Van Aken (based on Bunge) defines a technological rule as “...an instruction to perform a finite number of acts in a given order and with a given aim.” (Bunge, 1967); and a technological rule is “a chunk of general knowledge, linking an intervention or artefact with a desired outcome or performance in a certain field of application”: (van Aken, 2004).

A technological rule is general, which for IS design knowledge means that a rule is a general prescription for a class of IS-problems and not a specific prescription for a specific situation (for the X-IS-intervention in organization Z). Since a technological rule should be used by practitioner it should be applicable and actionable. Generally, the form of the technological rules is like “if you want to achieve A (outcome) in situation B (problem) and C (context), then something like action/intervention D can help because E (reason)”. Something like action/intervention D means that the rule is to be used as a design exemplar.

A field-tested and grounded technological rule has been tested empirically and is grounded in science. The latter means primarily grounding in results and theories from the behavioral
science paradigm. How to develop and test technological rules will be presented in Section 4.4. Field-tested and grounded technological rules will in most cases be in the form of heuristics. This is consistent with critical realism and means that the indeterminate nature of a heuristic technological rule makes it impossible to prove its effects conclusively, but it can be tested in context, which in turn can lead to sufficient supporting evidence.

4.4 Developing IS design knowledge

Van Aken (2004) suggests that management theory design science has much in common with evaluation research of social programs based on the philosophy of critical realism. We agree entirely with van Aken and suggest that evaluation research based on CR can make a major contribution to IS design science research. Related work has started on developing a critical realistic IS evaluation perspective (Carlsson, 2003) which builds on critical realism and realistic evaluation (Pawson and Tilley, 1997; Kazi, 2003; Mark et al., 2000). In our IS design science research framework the intention is to produce ever more detailed answers to the question of why and how an IS initiative works, for whom, and in what circumstances. Using the framework means that a researcher attends to how and why an IS initiative has the potential to cause the (desired) change. In this perspective, an IS design science (ISDS) researcher works as an experimental scientist, but not according to the logics of the traditional experimental evaluation research. Bhaskar states: “The experimental scientist must perform two essential functions in an experiment. First, he must trigger the mechanism under study to ensure that it is active; and secondly, he must prevent any interference with the operation of the mechanism. These activities could be designated as ‘experimental production’ and ‘experimental control’.” (Bhaskar 1998). Figure 2 depicts the realist experiment.

ISDS researchers orient their thinking to context-mechanism-outcome pattern configurations (CMO configurations). This leads to the development of transferable and cumulative lessons from ISDS research. A CMO configuration is a proposition stating what it is about an IS-initiative which works for whom in what circumstances. A refined CMO configuration is the finding of an evaluation of an IS initiative.

ISDS researchers examine outcome patterns in a theory-testing role. This means that an ISDS researcher tries to understand what the outcomes of an IS initiative are and how the outcomes are produced. Hence, the researcher does not just inspect outcomes in order to see if an IS-initiative works, but analyzes the outcomes to discover if the conjectured mechanism/context theories are confirmed.

In terms of generalization, an ISDS researcher through a process of CMO configuration abstraction creates “middle-range” theories. These theories provide analytical frameworks for interpreting differences and similarities between classes of IS-initiatives. Given that the goal is to develop design theories and knowledge—to construct and test context-mechanism-outcome pattern explanations—for practitioners ISDS researchers need to engage in a teacher-learner relationship with IS practitioners.

ISDS research employs no standard research design formula. The base strategy is to develop a clear theory of IS initiative mechanisms, contexts and outcomes. Given the base strategy, an ISDS researcher has to design appropriate empirical methods, measures, and comparisons. ISDS research is supportive of the use of both quantitative and qualitative methods.
ISDS research based on the above can be carried out through an IS design science research cycle (Figure 3). The starting point is theory. Theory includes propositions on how the mechanisms introduced by an IS-invention into a pre-existing context can generate outcomes. This entails theoretical analysis of mechanisms, contexts, and expected outcomes. This is the first step in developing technological rules and means that one tries to generate technological rules using our current knowledge, that is, grounding in theory. The second step consists of generating more specific “hypotheses”. Typically the following questions would be addressed in the hypotheses: 1) what changes or outcomes will be brought about by an IS-intervention (initiative), 2) what contexts impinge on this, and 3) what mechanisms (social, cultural and others) would enable these changes, and which one may disable the intervention. In this step the technological rules are refined.

Figure 3: The Information Systems design science research cycle—based on Pawson and Tilley (1997) and Kazi (2003)

The third step is the empirical test and includes selection of appropriate data collection methods. ISDS research is supportive of: 1) the use of both quantitative and qualitative evaluation methods, 2) the use of extensive and intensive research design, and 3) the use of fixed and flexible research design. In this step it might be possible to generate evidence of the IS-intervention’s ability to change reality. Based on the result from the third step, we may return to the IS-intervention to make it more specific as an intervention of practice. Next, but not finally, we return to theory. The theory may be developed, the hypotheses and the technological rules refined, the data collection methods enhanced, etc. To develop the technological rules means that the cycle will be repeated. As said above most of the technological rules will be heuristic. Through multiple case-studies one can accumulate supporting evidence which can continue until ‘theoretical saturation’ has been obtained.

5. Conclusion and further research

This paper points out some limitations and weaknesses in the current IS design science research frameworks and suggests that critical realism (CR) could be a fruitful philosophical underpinning for IS design science research and an IS design science framework. We presented a framework based in critical realism and having a broader view on what types of knowledge IS design science research should produce. This broader view is a direct consequence of that we do not just focus the IT artefact, but instead focus IS. Our framework can be summarized as (adapted from van Aken, 2004):

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>IS design science research framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant paradigm</td>
<td>Design sciences</td>
</tr>
<tr>
<td>Focus</td>
<td>Solution focused</td>
</tr>
<tr>
<td>Perspective</td>
<td>Researcher as experimenter (player)</td>
</tr>
<tr>
<td>Logic</td>
<td>Intervention-outcome</td>
</tr>
<tr>
<td>Typical research question</td>
<td>Alternative IS interventions for a class of problems</td>
</tr>
<tr>
<td>Typical research product</td>
<td>Tested and grounded technological rules (design knowledge)</td>
</tr>
<tr>
<td>Nature of research product</td>
<td>Heuristic</td>
</tr>
<tr>
<td>Justification</td>
<td>Saturated evidence</td>
</tr>
<tr>
<td>Type of resulting theory</td>
<td>Practical and abstract IS design theory and knowledge</td>
</tr>
</tbody>
</table>

Further theoretical and empirical work is required to develop and test the use of CR in IS design science research. Our suggestions make no claims to be the final word in the debate on IS design science research, but research based on the framework can lead to a stream of research that can develop high scholarly quality and practical (professional) IS design science knowledge.

References


A Generic Toolkit for the Successful Management of Delphi Studies

Jacqueline Day and Milena Bobeva
Business School, Institute of Business and Law, Bournemouth University, UK
jday@bournemouth.ac.uk
mbobeva@bournemouth.ac.uk

Abstract: This paper presents the case of a non-traditional use of the Delphi method for theory evaluation. On the basis of experience gained through secondary and primary research, a generic decision toolkit for Delphi studies is proposed, comprising of taxonomy of Delphi design choices, a stage model and critical methodological decisions. These research tools will help to increase confidence when adopting the Delphi alternative and allow for a wider and more comprehensive recognition of the method within both scientific and interpretivist studies.

Keywords: Research method, Delphi, Research Design, Research Evaluation

1. Introduction

Delphi is a structured group communication method for soliciting expert opinion about complex problems or novel ideas, through the use of a series of questionnaires and controlled feedback. Delphi has been well explored in a variety of areas, including government, medical, environmental and social studies, as well as business and industrial research (Linstone and Turoff, 2002), but had limited use in Information Systems (IS) research (e.g. Branchseau et al. 1996; Galliers et al. 1994, Schmidt et al. 2001). In all subject domains Delphi has been primarily employed for forecasting, planning, issue identification/prioritisation, or for framework/strategies development (Okoli and Pawlowski, 2004). Thus the method has mainly been used for theory generation, rather than testing and evaluation (Holsapple and Joshi, 2002). Furthermore, particularly in the IS research, the analysis of the Delphi has concentrated upon reporting the results of the study: there has been limited reflection and evaluation upon the use of the research method itself. To address this gap, the present work provides details of the design and application of the Delphi for empirical evaluation of a framework for Information Architecture (IA) for business networks.

This research builds upon previous work that argues for wider and more comprehensive recognition of the Delphi seeing it as a method that supports either scientific or interpretivistic studies (Day and Bobeva, 2004). The paper also seeks to further address the gap in research methodology by consolidating the authors’ experience with Delphi into a toolkit that includes generic stage and critical decision factor models (Fig.1). The purpose of this toolkit is to enable the viability and appropriateness of a Delphi inquiry to be established and to identify the practical limits upon its use. Key decisions have been identified to help guide the conduct of an inquiry, by isolating those decisions that have the most influence upon the standard of the final research deliverables. If substantiated as a well-proven and robust set of communication processes, the Delphi enquiry could gain wider use for many, if not most, studies of the managerial and business aspects of information systems and technologies.

2. Characteristics of the Delphi method

The Delphi is founded upon the use of techniques that aim to develop from a group of informants an agreed view or shared interpretation of an emerging topic area or subject for which there is contradiction or indeed controversy. Delphi shares that part of the research method continuum with several other consensus-seeking approaches. The family of methods have been mapped within Figure 2, which shows a simple arrangement, reflecting the relative degree to which the researcher and informants are interactively engaged with each other during the course of the research process. Most of the ways for acquiring data or evidence will be quite familiar...
to business researchers, except perhaps Nominal Group Technique (NGT). This is closely related to brainstorming but follows a highly structured agenda for building the base of knowledge through the use of formal means to involve all participants. An example of NGT is information systems requirements definition by the use of Joint Application Development workshops (Andersen 2000).

![Diagram of brainstorming, NGT, and Delphi groups with varying intensities of communication.](image)

**Figure 2:** A classification of consensus seeking methods

Although Delphi emerged in the 1950s in a project with strategic importance and was first reported in 1964, it could be argued that in comparison with the usual survey, Delphi is still in its developmental stage. However, the method has gained recognition mostly amongst communities dealing with complex problems, where Linstone (1978) argues that Delphi can be invoked as a method of 'last resort', being particularly suitable for situations where:

- the gathering of subjective judgements moderated through group consensus is the only approach possible in the absence of precise analytical techniques, or
- personal contact is not possible due to time and cost constraints, or is not desirable, due to concerns about the difficulty of ensuring democratic participation.

Whatever the perceived reasons for its choice, the method offers reliability and generalisability of outcomes, ensured through iteration of rounds for data collection and analysis, guided by the principles of democratic participation and anonymity. However, Turoff and Hiltz (1996) warn that because of its emphasis upon communication, Delphi can be in danger of dismissal as merely a form of data collection, when it is much more than this. Its iterative feedback method develops an insight, which in its totality, is more than the sum of the parts.

A problem with Delphi is that it is difficult to draft an explicit all-encompassing definition of the method. Here the present authors concur with Linstone and Turoff (2002). A preliminary taxonomy of Delphi design variations demonstrates that in addition to subject domain criterion the wide spectrum of Delphi applications may be categorised in terms of the following:

- **Purpose of the study:** building, exploration, testing, evaluation. The prevalence of studies focusing on developing and exploration has been discussed at the beginning of the paper.
- **Number of rounds:** varying between two and ten (Lang, 1994; Errfmeyer et al, 1986) but most commonly restricted to two or three rounds. Gottschalk (2000), however, in his comparison of methodological choices identifies Delphi studies with only one round.
- **Participants:** homogeneous or heterogeneous groups. The profile of the participants could be defined by age, nationality, knowledge, expertise, qualifications, occupation or position and thus could be used to further differentiate between two applications of the method. Of particular importance to potential users of Delphi is establishing the expertise of the participant (Gordon, 1994) that affects the quality of the outcomes.
- **Mode:** face-to-face discussion or remote access. This classification is linked to the anonymity of the participants. Participation through postal or electronic communications allows ensuring full anonymity of the informants.
- **Anonymity:** full or partial. This was a key element of the original Delphi process. It is a principle that can be sustained even with face-to-face contact, if the study is appropriately designed. Although with these cases it is not possible to ensure full anonymity, as the participants will know each other, their contributions to the study can remain anonymous. Similar considerations apply to technology-based Delphi implementations (see below).
- **Media:** paper-and-pen based, through telephone/fax, or computerised. Developments in information and communication technologies (ICT) have stimulated and been driven by changes in business and society. The convenience of electronic communication has steered the evolution of the Delphi toward computer-mediated studies. This could foster further developments, including support from multimedia, simulation and modelling tools and altogether boost new research opportunities for the method:
"When this technical capability is coupled to the knowledge being gained in the area of Delphi design, all sorts of opportunities seems to present themselves."

(Linstone and Turoff, 2002, p. 483)

- Concurrency: classic sequential set of rounds or real-time online conferencing. A consequence of ICT, such as video conferencing, is opportunities to vary concurrency modes, depending upon the nature of the problem and the urgency of its resolution.

This rich menu of Delphi applications could be developed further through designs based upon the different combination of types (Table 1), which forms the tool for Delphi studies:

Table 1: Taxonomy of Delphi Inquiry designs

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose of the study</td>
<td>building, exploration, testing, evaluation</td>
</tr>
<tr>
<td>Number of rounds</td>
<td>between two and ten</td>
</tr>
<tr>
<td>Participants</td>
<td>homogeneous or heterogeneous groups</td>
</tr>
<tr>
<td>Mode of operation</td>
<td>face-to-face or remote access</td>
</tr>
<tr>
<td>Anonymity of panel</td>
<td>full or partial</td>
</tr>
<tr>
<td>Communication media</td>
<td>paper-and-pen based, through telephone/fax facilitated, computerised</td>
</tr>
<tr>
<td>Concurrency of rounds</td>
<td>sequential set of rounds or real-time online conferencing</td>
</tr>
</tbody>
</table>

3. The Delphi Stage model

The second component of the toolkit is best understood through a detailed examination of an example. The empirical case chosen to illustrate the Delphi Stage model is work that was done to test the desirability and feasibility of an information Architecture to support e-business alliances. Although the example is, of course, specific to a particular variant of the method, the model itself is grounded in Delphi research undertaken by many authors. Generalisation allows the model to be used for all Delphi inquiries, regardless of its subject field or purpose, including the building, exploration or evaluation of strategies, policies, theories, etc.

3.1 The Delphi test for an Information Architecture framework

Information Architecture could be defined as a blueprint for principles, guidelines, standards and models for information requirements, systems development and management. It has long been seen as major issue for IS managers (Galliers et al, 1994, Galliers, 1995; Pavlia and Wang, 1995; Pavlia et al., 2002; Watson and Branchau, 1992). Although the spectrum of research topics is generally very rich, to date, little attention has been paid to specific architectures for electronically integrated businesses. The work carried under the banner of IA has either been constrained by organisational/sector boundaries or the conflation of web architectures with information architectures (Rosenfeld and Morville, 1998). Furthermore, since e-mediated business networks are at an embryonic stage, research was needed to sense the opinion of experts in this field and challenge them to achieve a consensus about the key components of a theoretical framework for an e-business IA.

Delphi was thought to be suitable because it would enable a clear epistemological stand to be taken in this study, i.e. the post-positivist position that enforces the merits of the scientific inquiry with interpretivist features. The fact that Delphi has not been fully exploited for defining the characteristics of complex entities, such as IS/IT architectures (Mulligan, 2002), neither for theory testing/evaluation, was also a strong incentive for testing the feasibility of the method.

Apart from the constraints imposed by the goal of evaluating the desirability and feasibility of the IA framework, the study were influenced by the specifics of both the research population, which was geographically distributed and a high degree of reliance placed upon electronic communications with information specialists. To maximise the quality of the outcome and address concerns for methodological rigour, the Delphi study was triangulated with a parallel electronic survey and follow-up evaluation interviews. The survey employed a similar data collection instrument and was used as a control group. However, the interviews and their use for evaluating the quality of the IA framework are beyond the scope of this paper. An example showing the use of design choice selection tool is presented as Table 2.

Table 2: The Delphi design choices in the study of Information Architecture

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Choice for IA Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose of the study</td>
<td>testing</td>
</tr>
<tr>
<td>Number of rounds</td>
<td>three</td>
</tr>
<tr>
<td>Participants</td>
<td>heterogeneous group</td>
</tr>
<tr>
<td>Mode of operation</td>
<td>remote</td>
</tr>
<tr>
<td>Anonymity of panel</td>
<td>full</td>
</tr>
<tr>
<td>Communication media</td>
<td>paper-and-pen based</td>
</tr>
<tr>
<td>Concurrency of rounds</td>
<td>sequential</td>
</tr>
</tbody>
</table>
3.2 Key stages in a Delphi study

An analysis of the process followed and the relevant literature yielded a generic Delphi model, comprising three stages: Exploration, Distillation and Utilisation. Details of the activities and results of each stage of the empirical research are shown in Figure 2 and discussed below.

The first stage, called ‘Exploration’ (Linstone and Turoff, 2002; Ziglio, 1996) is a free-flowing and unstructured investigation of the issues, limitations, challenges and problems that affect or are affected by the elements within the study domain. It includes the following activities:

- establishing criteria for selection of participants
- establishment of a Delphi panel
- design of the data collection and analysis instruments
- eliciting the initial set of issues to be tested through the Delphi rounds, and
- piloting of the toolkit.

Previous Delphi research has recognised that preparatory effort is necessary before the start of the rounds, but does not distinguish this work as a separate stage. Furthermore, the development of the initial base of knowledge has been traditionally considered as a first round of the study. The starting position for the Delphi (the first version of the questions) can be established by either exploratory or confirmatory in nature. The former is best implemented by seeking the views of the informants through initial open-ended questionnaires. The latter is appropriate for the more vague, ill-defined or contradictory situations often found in social, political and/or organisational worlds. It emphasises the qualitative dimension to the Delphi and hints that it might be successfully integrated with other research methods (see Section 4). The confirmatory form of the Delphi initial stage is traditionally carried out by circulating a predefined list of issues to the panel (Niederman et al. 1991). This is typical mode of working for follow up studies (Brancheau et al. 1996, Gottschalk 2000). For the present work the introduction of the stage-organisation allows for the initial round to be distinguished from the rest of the Delphi iterations, based upon the differences in goals: that is, ‘generation’ vs. ‘evaluation and extension’. This refinement allows for further development of Delphi variants, allowing application for theory testing and extension, where the theory is generated through either secondary or primary research. In this particular application of the method for evaluation of a conceptual IA framework, the list of issues presented to the participants reflected the perspectives and components of the proposed framework, thus negating the need for an initial ‘generation’ round.

This modification to Delphi has been employed in other studies (Custer et al. 1999; Doke and Swanson, 1995; Mulligan, 2002). Its merits are seen primarily for addressing the weaknesses arising from the unstructured polling of participants’ views, by replacing this initial collection of opinions with synthesis of key issues identified in the literature or through preliminary interviews with selected domain experts.

The remainder of the present study was focused on developing a group consensus about the architectural framework components and adhered to the normal form of inquiry for a ‘ranking type’ Delphi (Schmidt, 1997). This included a series of three rounds, conducted over a period of twelve months. After collecting the participants’ scores, a convergence ratio was determined. This represents the extent of participant agreement about the ranking of the architectural elements. Measuring the percentage of votes that fall within a prescribed range is a common approach to assess consensus. However, Scheibe, Skutsch and Schofer (1975), cited in Linstone (1978), argue that this is not the most reliable of measures and accordingly two further tests were implemented:

- Stability: monitoring the permanence of respondents’ vote distribution over successive rounds. This is idea is based upon Linstone’s (1978) view that opinion stability reflects consensus. He suggests that marginal changes of less than 15% offer a working definition of a threshold for stability, which might be used as a criterion for termination of the Delphi exercise.

- Participation: whether the numbers of participants will drop below a critical level. Ziglio (1996) asserts that useful results can be obtained from small size, homogeneous groups of 10-15 experts; whilst Dalkey, cited in Linstone (1978, p.296), found that seven is a suitable minimum panel size. This was a useful measure in a case such as this study, where the drop in response rates was about 40% after at each round.
Delphi Planning
- Transposing the framework into a set of questions
- Formation of the set of criteria for participants selection
- Preparing the set of questionnaires and supporting letters

Selection of participants
- Arranging the dispatch of the first/second batch of Round1 questionnaire - collaborative work with the Placement office
- 90 IS professionals selected for the first batch of Round 1
- Survey with BISM Year 4 students to identify potential participants in the Delphi study and/or the follow up interviews
- 69 IS professionals selected for the second batch (4 addresses not found, figure for the second batch - 65 effectively).
- Round1 questionnaire sent to 159 participants in total (135 effectively, as 4 addresses were not found).

Pilot study
- Conduct 3 pilots of the questionnaire
- Amend the design and content as per the recommendations of the participants in the pilot

Delphi Round 1
- Sample size: 185
- Objective: To test the desirability and feasibility of components of the framework
- Contents of the survey: framework components (for each issue: desirability and feasibility mark on a Likert scale of 1 to 10 and space for comments), organisational specifics with regards to information exchange, personal information
- Monitoring the return rate. Less than 11% at the end of the first month. Telephone reminders.
- Minor modifications done to the design of the questionnaire for the first batch based on recommendations from one participant.
- Return rate from the first batch: 16.67% (15 out of 90); Return rate from the second batch: 6.15% (4 out of 65)
- Return rate total for Round 1: 12.26%
- Total number of responses: 19 out of 155 (1 anonymous)

Progress to Delphi Round 2 or Termination of the Delphi survey
- Review of the returns from Round 1.
- 18 is considered as acceptable sample size; Proceed to Round 2.

Delphi Round 2
- Sample size: 18
- Objective: To inform the participants of the results from Round1 and give then an opportunity to review their scorings in the light of the average results
- Contents of the survey: average result, participant score from the Round1, amendments, comments
- Reminder sent via e-mail ten days after the second questionnaire was sent.
- Return rate: 66.67% (12 out of 18)

Progress to Delphi Round 3 or Termination of the Delphi survey
- Analysis of the results from Round 2 with regards to changes of opinion.
- 9.76% (15.3%) shows stability in the respondents’ views. Proceed to Round 3.

Delphi Round 3
- Sample size: 12
- Objective: To inform the participants of the results from Round2 and give then an opportunity to review their scorings in the light of the average results
- Contents of the survey: average result, participant score from the Round2, amendments, comments
- One questionnaire returned as the participant had left the company.
- Return rate: 63.6% (7 out of 11)

Progress to Delphi Round 3 or Termination of the Delphi survey
- Analysis of the results from Round 3 with regards to changes of opinion and size of the sample.
- Decision to terminate the study due to the small size of the sample.

Analysis of the Delphi study results
- Analysis of the results in categories of desirability and feasibility
- Analysis of the significance of the new components the IA framework introduces
- Consideration and analysis of the qualitative feedback, i.e. the comments of the participants
- Analysis of the impact of the results of the survey on the proposed framework

Utilisation of the Delphi results & experience
- Drafting the summary of the findings from the Delphi study, that is to be sent to those participants that indicated that they would be interested in the results.
- Using the results of and the experience with the Delphi study when designing the on-line survey and reviewing the FEBuS framework.
- Reflecting on the experience gained from the application of the Delphi study.

Figure 2: Delphi study implementation process
Repeated attempts at opinion-seeking and subsequent analysis to see whether or not the Delphi had reached a critical point for study termination, forms the Distillation stage in the generic Delphi model. It is a distinctive feature of this study that the three tests were applied at each round. Anecdotal evidence from Delphi studies across various disciplines suggests that few inquiries use more than one of these tests.

To be consistent, the first two stages have been named in accordance with the terminology devised by Ziglio (1996), although with non-similar content. The last stage, Utilization, has not been recognized as distinct until now. The present paper proposes that the activities within this stage be classed as 'short' or 'long' term in nature. The former includes the development and dissemination of the final report on the study. Long-term activity is knowledge dissemination about the Delphi exercise.

4. Some critical issues in Delphi inquiries.

The third and last component of the generic Delphi toolkit is a checklist that can be used to ensure that key operational decisions about the management of a Delphi study have been thought through. A link may be made with the stage model discussed in Section 2.2. The researcher's knowledge, ideals, awareness, flexibility and openness to feedback will affect the extent to which these key decisions are able to be fully addressed (Fig.3).

![Critical Issues Sets](image)

**Figure 3:** Critical methodological issues in a Delphi inquiry

### 4.1 Design constraints

A Delphi study, even if simplified as an integrated series of surveys, each progressively building upon findings, requires more time and effort than many of the methodological alternatives. As such, it is unlikely to be helpful where knowledge can be relatively easily and fully acquired through direct measurement, experiment or simulation. Thus a clear justification for the choice of a Delphi inquiry, are the characteristics of the participants, sample size and design of the data collection instruments. These are the first critical decisions a researcher will face.

#### 4.1.1 Critical issue 1: Choosing the approach

Given that the theoretical foundations of an inquiry, such as the epistemological position and research axioms have already been established, an early decision that confronts the researcher is the choice of an approach that meets the requirements of the research philosophy.

For positivist research, Delphi can be used for defining the characteristics of complex areas (such as IT strategy) and for testing general propositions (nomothetic perspectives), including those concerning the attitudes of social actors. The latter was the case with the IA research that has been used as the exemplar. Since Delphi inquiries are anchored in aggregations of opinion, they are not helpful for investigating psychosocial conditions of an individual, so Delphi is not recommended for research where nuances and experiences of human individual behaviours must be studied in situ. Here evidence must necessarily be gathered through observation of people and the context for their actions. However, since it does not disturb the naturalness of the setting, Delphi can support some types of field studies. The iterative cycle of Delphi is supportive of longitudinal investigations. For example, theory development directed toward building a process model from a factor-based framework.

#### 4.1.2 Critical issue 2: Initial Selection of the Panel

A Delphi panel should consist of individuals with knowledge about the substantive area of research, the motivation to engage with the inquiry process and be able to articulate judgements. The first problem with the selection of informants is when there is no clear definable community that can act as source of expertise, knowledge or opinion. This is often the case of future studies, where the existing knowledge base is of limited value for predicting the future and is best served by constituting the panel from experts drawn from
A wide range of subject areas. Similarly, there can be difficulties where informants are identifiable or the group is so homogenous and constrained by conditions that diversity of opinions cannot be represented sufficiently strongly to drive the goal-seeking heuristic of Delphi. Here, the researcher might be in a quandary as to the attitudes of the ‘expert’ to be recruited to the panel: should they hold strong, positive opinions or is it better for them to be well-balanced and impartial? Goodman (1987) in Hasson et al (2000) argues that individuals are more likely to get involved in the Delphi inquiry if they are to be affected directly and/or profoundly by the outcome of the study. Thus for political and medical topics studied neutrality is likely to be less useful than for business management or IS research. It is the nature of Delphi to ‘trim’ the extreme cases but we argue later that in all situations these extreme or maverick opinions should be carefully conserved because they can act as compare-and-contrast case for establishing the validity and generalisability of consensus opinion.

If there is a large potential population of experts then the researcher must decide the basis on which the sample is to be chosen to form the panel. The choice remains, as for other forms of research, between probability and non-probability (purposive) sampling (with all the possible variations possible within the latter). In practice, the strategy selected is likely to depend upon the nature of the research problem: the narrower the scope, the greater the depth and specificity of expertise needed and the more likely a purposive approach is appropriate. On the other hand, for wide-ranging, social and marketing Delphi studies, an initial ‘random’ sampling can be used.

The sampling regime that is implemented will be also be influenced by the form of the initial contact made with participants: cold calling or after some groundwork has been done. It is probable that a more personal approach will result in a positive response to an entreaty to enrol as a panellist. However, no matter what initial sampling strategy is adopted for all types of Delphi study, the research outcomes will be derived form data collected from a self-selected sub-set of the original population.

In the empirical IA study, panelists were selected through non-probability sampling. More than half of the respondents were IS project managers with widely differing levels of experience. Other individuals were employed in managerial positions or as consultants in systems analysis, data architecture, outsourcing or sales. Dalkey, Brown and Cochran (1969), cited in Linstone (1978, p.296), argued that self-rating of expertise in the area of research can be used to improve the efficacy of a Delphi inquiry.

Finally, with regard to the constitution of the expert group, it is important to recognise that the Delphi researcher is not immune to implementing good general practice, common courtesies and ethical standards. The principle to be followed here is meeting expectations. For example, clearly informing participants upfront of what is expected from them and when and by the researcher ensuring that promises made to panellists concerning their anonymity or within the subsequent dissemination of the study results are kept.

4.1.3 Critical issue 3: Ongoing Management of the Panel

The drop out rate for a Delphi study could be high after the first few cycles thereby reducing the sample population for subsequent statistical evaluation and indeed this is affect was experienced by this study. The size of the panel that is needed at each round is a decision that must be taken before the execution of each these iterations. Although the largest Delphi undertaken in Japan involved several thousand people (Linstone 1978), most studies use panels of between 15 to 35 people (Gordon, 1994), though Dalkey (2001) suggests that seven as the minimum number. Once the requirements for participant expertise (Issue 2) have been clarified, different sampling strategies can be used to constitute the panel. In the IA project, 155 IS professionals were approached, of which 19 (12%) agreed to take part in the initial Delphi round.

4.1.4 Critical issue 4: Instrument design

The design of the data collection instrument is critical for both the Exploration and Distillation stages. Creativity is always a good watchword for any data collection mechanism, so different ways of conceptualising the topic and structuring the questions should be tried to match communication mode with the individual preferences of the informants. The design of the instrument will obviously depend a good deal upon the number of questions asked. There are no clear rules for this. Hasson et al (2000) suggest that the minimum number of issues should be six, but this does not exclude the possibility of there being more, e.g. ten
(Niederman et al 1991) or 34 as is the case for the present study. The number issues explored will reflect the complexity of the problem and to the type of data collected.

Further, the researcher should be aware of structuring the questions in a way that implies an answer or does not properly allow for different views or for a novel re-framing of the problem. The results of the present study also bear out well-proven ideas about survey design. The key is formulating clear, concise and unambiguous questions, together with clear instructions for the participants. Experience supports the common sense view that aesthetically pleasing and easy to fill in questionnaires positively influences an informant’s decision whether or not to take part a study.

For the IA study a 10-point Likert scale was used, where ‘one’ coded the least desirable/feasible constituent and ‘ten’ for the most desirable/feasible IA component (see Figure 4 and 5 for example of questions and layout). The flow of the questions deliberately did not reflect the logical organisation of the framework, with some elements assigned a higher priority than were others. Rather, questions were grouped by perceived simplicity of the subject matter, starting with questions about the more straightforward topics of IA. Space for comments and/or clarifications was also provided adjacent to each answer block. On completion of the core part of the instrument, further questions were then posed about information exchange, so as to probe the extent to which the participant’s organisation was a true member of an e-business network. This was also done to identify appropriate informants for the follow-up interviews. The Delphi survey instrument concluded by asking personal information, such as contact details and preferred method for communication, as well as ascertaining whether the panellist would be interested in the results from the completed study.

![Figure 4: Questionnaire layout for Round One (Page 1 of 3 shown only).]
4.2 Implementation factors

In addition to the instrument design and sample size there are a few other decision points throughout the implementation of the Delphi rounds. These are discussed below.

4.2.1 Critical issue 5: Timing of the Delphi survey

This issue was encountered within the IA study and it can have a significant impact upon the stability of the results and the response rate. Greater volatility of the participant scores was observed in the third, compared to the second Delphi round. This could be attributed to the time that elapsed (ten months) between the second and the third iterations, compared to two months between Rounds One and Two. The wider time gap allowed for more change in an individual's circumstances, knowledge and situational context. Although the period chosen for dispatch of the questionnaire was designed for the maximum availability of potential respondents (and thus the highest expected response rate) the first round rate was 12%, and a third of the participants opted out of the study at each of the next two rounds. This limited response may be due to 'survey fatigue': perhaps, nowadays there are too many surveys and some antipathy arising from the failed promises of e-business!

4.2.2 Critical issue 6: Proactive management of the study

This refers to the researcher's ability to maintain a high level of communication during the rounds to enable an adequate level of response to be obtained and to know when to terminate the study. The convergence, stability and participation tests outlined in Section 2.2 provide a good sense of the development of the inquiry after round and could be used as criteria for termination of the study. As mentioned earlier, anonymity is a key feature of Delphi studies, though computer-mediation can enable conversations between informants to take place and may help panel members to more readily consider acceptance of a wider diversity of views. Complementing the traditional version with conferencing features may add significant advantages but presents a challenge to the skilful management of Delphi, because the researcher must be able to facilitate on-line discussions.

4.2.3 Critical issue 7: Documentation of the results

In addition to the obvious task of clearly documenting the results of all rounds, the authors argue that documentation of results includes consistently recording divergent views at a similar level of detail. It is also important that informants are encouraged to provide reasons for the change of views and these also should be captured as part of the
base of data. Restrainted use of graphical presentation of the development of the consensus (Malhotra et al. 1994) is recommended here not only to help the researcher manage the operation of the rounds but to enable informants to better locate their individual views within the consensus. By doing this ownership of the research is engendered and continued participation is maintained.

4.2.4 Critical issue 8: Analysis of the inquiry results

Problems can arise in cases where the data collection instruments employ an interval scale (such as Likert) to measure the importance of an item since a correct choice of statistical analysis methods must be made. The statistics included as part of reported Delphi studies mainly give details about significance in t-test. However these are parametric tests normally used with continuous variables, instead of non-parametric tests that should be used for analysing ordinal data. A standard way to present the Delphi results of a single iteration can be based upon the relative importance of the issues, as determined by the mean values for each element of the instrument. Both descriptive statistics and non-parametric tests were employed in the IA Delphi for exploring relationships and for comparing data groups. These analyses included a Friedman test of the changes to desirability and feasibility scores across the three rounds. Mann-Whitney ‘U’ tests were also used to examine the difference between the set of scores on feasibility and desirability of Delphi study participants compared to those in e-survey. A closer examination of the change patterns of the responses revealed that there were four extreme cases, one with an exceptionally high rate of changes and three with no change. Other Delphi studies were consulted for guidance about how to treat extreme results, but the literature provided no details about the change patterns but only change rates.

There was also a need to analyse qualitative evidence, so biases of the participants and researchers could be properly acknowledged. A difficulty with this is that at the moment there are few tools available for processing a large number of non-numerical, unstructured, and rich data sets that can be captured with in Delphi studies.

4.3 Evaluation criteria

The final set of decisions is concerned with making sense of the results obtained from a Delphi inquiry and ascertaining the quality of the final outcome.

4.3.1 Critical issue 9: Reviewing results

Given the claim that Delphi can span the divide between the positivist/quantitative and interpretative/qualitative ideals (Day and Bobeva, 2004), then range of evaluative perspectives partaking of both traditions is needed to review the quality of the Delphi findings. The trustworthiness criteria of confirmability, credibility, transferability and dependability could complement or replace the positivist criteria of objectivity, validity and reliability.

Confidence levels

In the exemplar the researchers acted purely as facilitators and not participants, but were aware of the threat that their subjective interpretation could have when reflecting upon changes in and the motives for the results from each round. Failure to understand the context for the consensus may lead to subsequent failure to capture important contextual information. In the absence of self-assessment tests for individuals forming the panel, systematic biases can arise from the judgements made by the participants reflecting the heuristics employed. A common problem with any non-administered survey is that psychological factors, causing random and systematic errors to impact upon the study and which may be difficult to detect. These ‘random’ errors can be attributed to factors such as work pressures, the time of the day when the survey was completed, the mood of the informant, as well as characteristics of the instrument itself, such as unclear instructions or ambiguous questions.

To enable confidence to be placed in the primary data, in addition to the criteria for reliability and validity of questionnaires suggested by Mitchell (1996), a detailed rationale should be provided showing how the question set was initially established and employed. This could be accompanied by an audit trail providing a chain of evidence through the data analysis stages.

Rigour

When evaluating the rigour of the research findings the criteria of internal validity could be strengthened by examining plausibility and consistency of the participation results of the panel members. Delphi implements these tests of quality through the continuous
feedback and confirmation by the contributors but researchers must exercise care in the execution of the rounds those important contextual changes are detected and properly acknowledged.

To maximise confidence in the example project, the formulation of the architectural components was based on existing models and theories. Evidence was accumulated to confirm the most common components, derived from these existing architectures and how they should be represented in the proposed framework. The elaboration and comparison strategy proposed by Reichardt (2000) was then used to improve development by the reporting of average scores for each construct in the framework. However, a significant issue for the present research was the ability of the questionnaire to properly represent the proposed framework, as in some instances, a single question addressed more than one information category. Usually this reflected a relationship between the data categories.

Credibility

Lee (1999) rightly observes that the use of multiple informants helps to strengthen credibility. For the IA study this meant comparing the results derived from one part of the research population with another. That is, by testing the views of the panellists against the on-line participants and interviewees to support a form of ‘triangulation.’ Minor variations found between the groups could be well attributed to the different contextual settings of individuals, as well as the capability of each person to understand the scope and objectives of the study. The strategy adopted to reduce this threat to plausibility was the inclusion of reasons for the investigation and by providing the researcher’s contact details for further clarification.

According to Lincoln and Guba (1985) replication of outcomes from another context is an acid test for external validity but this non-meaningful for Delphi studies. Gordan (1994 p.1) explains why this is so:

“Because the number of respondents is usually small, Delphi’s do not (are not intended to) produce statistically meaningful results; in other words, the results by any panel predict the response of a larger population or even a different Delphi panel. They represent the synthesis of opinion of the particular group, no more, or less.”

For Delphi studies it is more appropriate and useful to take a qualitative perspective by examining the results of the Delphi for their cogency, relevant and plausibility, by identifying the explicit limitations upon transferability of the results to other contexts. This requires that questions must be devised about informant’s personal situations for the judgements offered by panel experts. This was not possible for this study since the different situations of each participant precluded testing the general applicability of the framework.

A summary of the critical issues discussed above is given in Table 3. In addition, Table 4 could be used as a reference for the set of issues pertaining to each Delphi stage and/or as a timeline of each issue throughout the Delphi stages.

Table 3: Critical issues in Delphi studies

<table>
<thead>
<tr>
<th>Design constraints</th>
<th>Implementation factors</th>
<th>Evaluation criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Choosing the approach</td>
<td>(5) Timing of the survey</td>
<td>(9) Reviewing results: Complementing qualitative with quantitative quality criteria.</td>
</tr>
<tr>
<td>(2) Informant selection</td>
<td>(6) Proactive management</td>
<td></td>
</tr>
<tr>
<td>(3) Sample size</td>
<td>(7) Documentation of results</td>
<td></td>
</tr>
<tr>
<td>(4) Instrument design</td>
<td>(8) Analysis of results</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Critical issues cross-referenced to Delphi process stage model

<table>
<thead>
<tr>
<th>Issues \ Delphi Stage</th>
<th>CI 1</th>
<th>CI 2</th>
<th>CI 3</th>
<th>CI 4</th>
<th>CI 5</th>
<th>CI 6</th>
<th>CI 7</th>
<th>CI 8</th>
<th>CI 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploration</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Distillation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Utilisation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

5. Further development of Delphi and of the toolkit

There are two significant trends to the environment for the conduct of Delphi studies that will require the continued development of the tool kit. The first and perhaps most obvious new element is the advent of personal electronic communication. Incorporating this into the picture means that it is possible to recognise some hybrids derived canonical forms depicted earlier. The importance of this is the recognition that Delphi does not have to be used as a single approach but can be used profitably with other techniques. Figure 6
shows attested examples in the literature of the mixed forms and also indicates (question mark) where there is an opportunity for future research and development of other multifaceted research methods that draw upon the strengths of several parent methods and different modes of communication. For example, the embedding of focus groups, brainstorming or interviews within the initial stage of Delphi panel formation or the use of computerized conferencing for real-time Delphi studies (Turoff and Hiltz 1996; Linstone and Turoff 2002).

The other major dynamic concerns the evolving social and organisational settings for group consensus seeking approaches in general, and for Delphi, in particular. Traditionally, the PESTE (Political, Economic, Social, Technical and Environmental) factors have thought to apply to business but they also are likely to impact upon the ways and means for conducting business and management research. This could include, for example, ‘extended’ Delphi studies, where the content of the feedback is enriched to include more anecdotal and qualitative information to enable a more informed decision to be made on the degree of consensus (Bobeva and Day 2005). There is much work to be done in reflecting legislation requirements (such as privacy) in research design and by exploring ways to tailor research processes to properly reflect international settings. For example, in some societies, open debate and consensus seeking is a pervasive cultural norm in others a more formal and authoritarian management style applies. In latter situation, maybe Delphi needs shifted towards NGT model with electronic mediation that enables individuals to contribute ideas without being worried about being seen as ‘stepping out of line.’

<table>
<thead>
<tr>
<th>Brainstorming</th>
<th>NGT</th>
<th>Focus Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brainstorming ?</td>
<td>NGT ? Computerized Delphi</td>
<td>Focus group (computer-based conferencing) ?</td>
</tr>
<tr>
<td>Group Interview</td>
<td>Extended Delphi Study</td>
<td>Traditional Delphi Study</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Participatory</td>
<td>Non-intrusive</td>
<td>Anonymous</td>
</tr>
</tbody>
</table>

**Figure 6:** New technologies and consensus seeking methods

6. Conclusions and recommendations

The toolkit discussed in this paper has been proposed as a means of managing a Delphi enquiry. The design of the three tools recognises the critical decisions in the conduct of the inquiry and identifies practical techniques that if implemented, should help to ensure a rigorous and valid study. Much work, however, remains to be done, particularly in the areas of communications, results analysis and in ways of monitoring and controlling research quality. Although the experience of using the Delphi has encouraged the authors to expand their personal research repertoire, generally the power of Delphi as an effective research method has unfortunately remained obscured through a lack of understanding. The contributions of other researchers are therefore vital to break through this conceptual barrier. Their efforts will be well-rewarded since they will acquire a flexible and simple way for exploring and evaluating many challenging topics in the realm of technological, managerial and organisation studies.

**References**


Learning Logs: Assessment or Research Method?

Tim Friesner¹ and Mike Hart²
¹Business and Management, University of Chichester, UK
²Business Management Group, University of Winchester, UK

t.friesner@ucc.ac.uk
mike.hart@winchester.ac.uk

Abstract: Learning logs are an increasingly popular mode assessment. They record learning, experience and reflection. This paper considers learning logs as a research method, where researchers wish to gain a deep understanding of the processes of learning, reflection and experience as they occur in individuals over a period of time. Techniques are offered for implementing logs as a research method, analysing the data and interpreting results.

Keywords: Learning logs, research method, reflection, learning, experiential learning, experience.

1. Introduction

A learning log is a vehicle that is used to assess learning from experience. Logs are an increasingly popular tool, often used in conjunction with work placements, work-based learning or courses that are underpinned by a philosophy that action learning is a pedagogical approach that best achieves learning outcomes. Learning logs are viewed firmly as an assessment. They are ideal for encouraging learners to reflect on learning, and they have a structure that is quite different to traditional assessments such as essays and reports. However they are also a source of reflective data. So for example, if one has 10 learning logs from 10 students that record learning over 10 weeks of work with 10 different companies, not only do you have 10 assessments, but also 10 case studies with very rich reflective data. From this perspective there is the potential to consider learning logs as not only an assessment but also as a research method. This paper evaluates the proposition that logs are a research method. It initially considers learning logs as an assessment and examines the nature and value of reflection. Then the structures of logs as both mode of assessment and research method are compared. An experiment using logs as a research method is described. Here data capture is discussed, and an integrated approach for interpreting the data is re-introduced (Friesner and Hart 2005). Finally the nature of reality and learning logs is examined, based upon Bannister 2005, before conclusions are made.

2. Learning logs as an assessment

Reflective assessment takes a number of forms. Cottrell (2003) summarises some of them as logs and journals, personal statements, position papers, reflective essays, progress files and portfolios, and presentations. This paper and the research experiment that it includes focus mainly on logs and journals, which for the purposes of this work are called learning logs or simply logs.

Learning logs are structured in many ways, often dependant upon the topic being studied, the level of a qualification, and the length of time over which the logging continues. Essentially there is usually an audit, for example a SWOT analysis or a skills analysis. The student then decides on which weaknesses to overcome, and which strengths to develop further. Objectives are set that conform to the acronym SMART (Specific, Measurable, Achievable, Realistic, and Timed). The logs themselves contain action points and plans that are designed to meet the objectives. Students begin their experience and record it in their logs. Logs can be done weekly, or when a significant event occurs. However, one must appreciate that learning logs are a formative assessment i.e. they are built overtime. To write a log at the end of an experience defeats one of its key uses, since logs capture cycles of learning that occur as a student develops. In each entry the student reflects on the extent to which their learning objective has been achieved. They may find that it has been achieved quickly, and then set a new, more substantial one. They may find that it has been partially achieved and set new action points. On the other hand, the student may find that the objective has not been achieved in any way, and have to change their action plan substantially, or modify the original objective. It is all part of the learning experience. At the end of the experience students often reflect on the whole learning process and the degree to which objectives have been realised. It is a synthetic reflection (Cottrell 2003) that allows a student to

Reference this paper as:
appreciate the linkages between their recent personal experiences and the bigger picture. Finally new objectives are set, in relation to life-long-learning and personal goals.

2.1 What is reflection?

Reflection is a type of thinking. It is associated with deep thought aimed at better understanding. (Cottrell 2003 p171)


Often a metaphor is used whereby an individual reflects on learning in the same way as they would reflect upon their physical presence using a mirror. To paraphrase Stella Cottrell’s popular view of reflection as an art, reflection helps us to make sense of learning from experience. It allows us to stand back and get a different view of experience. Reflection allows us to repeatedly go over an experience in our own mind, and allows for honesty. By undertaking reflection we weigh up aspects of our experiences and make judgements, and we may see our learning more clearly (i.e. in a mirror). It opens up the opportunity for deeper learning and understanding, allowing the learner to draw final conclusions.

Reflection can viewed as part of a more complex learning process in relation to action learning (Revans 1978) and experiential learning (Kolb 1984). For Revans (1978) action learning, or ‘Learning (L)’ was the addition of Programmed Knowledge (P) plus Questioning Insight (Q) i.e. L = P + Q, where Questioning Insight contains some of the properties of reflection.

Kolb (1984) offered a more detailed view of experiential learning, that includes reflection (or Reflective Observation) as an element of a more extensive learning cycle. Experiential learning is very much founded upon the Concept of Experiential Learning which explores the cyclical pattern of all learning from experience through reflection and conceptualising to action and on to further experience (Kolb 1984).

![Figure 1: The experiential learning cycle (Kolb 1984)](chart)

Which brings us back to learning log structure and how it collects data on reflection. The log itself represents a series of cycles that indicate that learning may or may not have taken place. On the other hand, by requiring a student to use a prescribed structure perhaps they are being forced through a learning cycle that is no real indicator of anything i.e. students are simply going through the motions. It is a recognised argument, and one that can be used to criticise learning logs as a mode of assessment in favour of a more traditional one such as an essay. However this point is not wholly accepted by this paper, since logs do have many benefits in terms of life long learning, as well as personal and professional development. Tables 1 and 2 below represent the close relationship between the structure or design of a learning log assessment in comparison with a learning log research method.

Table 1: Preparation Phase – a comparison between the application of a learning log structure to both assessment and research method

<table>
<thead>
<tr>
<th>Learning Log Assessment</th>
<th>Learning Log Research Method (Marketing Planning)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summarise your learning and development to date (e.g. from Personal Development Plan (PDP))</td>
<td>Summarise how you conduct marketing planning:</td>
</tr>
<tr>
<td>Summarise your current skills, strengths and weaknesses (in relation to subject under study)</td>
<td>Summarise your current marketing activity:</td>
</tr>
<tr>
<td>Complete a SWOT analysis</td>
<td>Complete a SWOT analysis</td>
</tr>
<tr>
<td>State personal aim(s) for the logging period</td>
<td>Decide upon strategies for achieving personal aim(s)</td>
</tr>
<tr>
<td>Decide upon strategies for achieving personal aim(s)</td>
<td>Decide upon strategies for achieving personal aim(s)</td>
</tr>
<tr>
<td>Summarise the organisation’s aim(s) for the next 12 months</td>
<td>Summarise the key marketing strategies of your organisation</td>
</tr>
<tr>
<td>Summarise the main marketing tactics of your organisation</td>
<td>Summarise the main marketing tactics of your organisation</td>
</tr>
</tbody>
</table>
Table 2: Action Plan Phase – a comparison between the application of a learning log structure to both assessment and research method

<table>
<thead>
<tr>
<th>Learning Log Assessment</th>
<th>Learning Log Research Method (Marketing Planning)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Reflective Journal considers: The aim(s) that you wish to achieve</td>
<td>The Marketing Planning Reflective Journal considers: The aim(s) that you wish to achieve</td>
</tr>
<tr>
<td>The Learning that takes place as log entries develop</td>
<td>Marketing planning and decision making</td>
</tr>
<tr>
<td>A series of reflective journal entries</td>
<td>Carry out and complete one journal entry per week, or on the occurrence of a ‘significant event.’</td>
</tr>
<tr>
<td>Each log entry should include:</td>
<td></td>
</tr>
<tr>
<td>The aim(s) of your organisation (from Preparation Phase above)</td>
<td></td>
</tr>
<tr>
<td>Action taken to achieve aim(s)</td>
<td></td>
</tr>
<tr>
<td>Evaluation of performance so far</td>
<td></td>
</tr>
<tr>
<td>Next steps (based on a, b and c)</td>
<td></td>
</tr>
</tbody>
</table>

Both approaches use audits at the outset. Objectives or aims are set and action points are put into place. Reflection is undertaken in log entries, and conclusions are drawn. Depending on the success or otherwise of the actions taken so far, new or adapted action points are put into place.

3. Learning logs as a research method

Having considered logs as an assessment, and then comparing a typical assessment structure with a log research method structure, the next section will consider the logging research method in more detail. An experiment is recounted that used a logging approach to gather the reflections of a group of marketing managers as they undertook a marketing planning process. Then consideration is given to the ways in which reflective data from logs can be analysed (Friesner and Hart 2005), and then nature of reality presented in logs is evaluated (Bannister 2005). Of course, any application of logs as a research method should take into account its suitability in comparison with other methods that may be better suited to the research problem faced or the questions posed. Really the research questions themselves should be seeking some reflection by individuals on learning and/or experience.

3.1 Research gathering data

An experiment was set up to help to answer research questions that were based upon learning, reflection and experience (Friesner and Hart 2005). Seven marketing managers from seven British theatres were recruited to participate in the learning log research experiment. Six ultimately completed. The research programme ran over an eight week period where marketing managers recorded marketing planning activities in a learning log that conformed to the same, pre-advised format. The research had two distinct phases (see Tables 1 and 2). Phase one was a preparation phase, and it was conducted prior to the commencement of logging. It required respondents to summarise their current marketing planning activities, by conducting a simple SWOT analysis, stating organisational aims for the next 12 month planning period, summarising strategies and tactics as they envisaged them over that period. This was a scene setting exercise that assisted participants to focus on the activity that lay before them. The second phase was the logging, or action phase. Each log entry began with the aim that the participant wanted to achieve in each weekly log entry. The entry continued with reflection upon the aim and action that had been undertaken to achieve it. The entry evaluated performance so far, and stated the next steps to realise the aim. This continued for 8 weeks, and then each marketing manager undertook a final reflection. The final reflection allowed respondents to reflect on not only the marketing planning that they had undertaken, but also their own individual learning and experience. The completed logs were 3,000 to 5,000 words. With the contextual data gathered at the outset of the experiment, the documentary evidence in its entirety closely resembled six case studies.

3.2 Analysing data from Learning logs

3.2.1 Content analysis, case study analysis and narrative and storytelling analysis

Since this kind of research had not been undertaken before, there was no precedent for ways in which the data should be analysed. A broad literature review was undertaken, and the most suitable approaches for analysing the data, in the view of the researchers, would be an amalgamation of content analysis, case study analysis, narrative and storytelling.
analysis, or some combination of the three (Friesner and Hart 2005).

3.2.3 Disadvantages of logs as a research method:

- A lot of effort and tenacity is needed to use logs as a research method.
- Replication and verifiability of findings is unlikely.
- Short, uncomplicated results are equally unlikely, since logs deliver rich data. Subtleties are not easily spotted in the data.
- There is an economic cost to managers that dedicate time to complete learning logs.

3.3 The nature of reality in learning log research

This section considers the nature of fact in interpretivist research. It is strongly influenced by a paper read to the 4th European Conference on Research Methods in Business and Management by Bannister (2005). The research experiment uses multiple case studies based upon surveys and learning log evidence, and issues relating to ‘reality’ and how it is represented by the data need some consideration. Bannister (2005) introduces the concepts of internal and then external reality, and the impact of a series of filters when examining multiple case studies. The external reality is what actually happened in the physical world, for example the timing of theatrical productions, or the way in which media was used to promote them. However internal reality or realities, represents the reflections of those respondents that completed the learning logs. The accounts are largely subjective and personal, and influenced by societal and political factors. Bannister (2005) stated that there are a number of predetermining filters, which he defines as mental processes through which facts and reality pass in moving from stage to stage in the research process. They are perceptual, contextual, linguistic, memory, sequence, personality, agenda, cognitive, methodological, and selection filters. It is recognised here that the ten filters are in evidence to a greater or lesser extent as the researcher interprets the data generated in the logs, and then goes on to consider findings and make conclusions. The researcher is reading the reflections of respondents based upon their perceptions of what happened. Perceptions may be loaded or biased, and may even be influenced by those around the marketing manager. Therefore, log entries tend to contain the perception of the authors, as they reflect upon events. The researcher is
also aware of contextual filters, since one has experience of marketing planning techniques. Therefore the researcher will be familiar with marketing planning vocabulary. Some respondents do not have a marketing planning vocabulary, and care should be taken to look for other evidence of the reality of marketing planning as the logs record it. The logs themselves contain the written word, and the researcher himself has to interpret the black and white page. The character of each individual, their confidence with the written word and indeed the quality of their written English may influence the way in which their personal reality is recorded. The way in which this researcher interprets the aforementioned words could distort the reality contained in the text. Indeed the interpretation of the accounts is then written up by this researcher, and again the nature of reality impacts upon the original intention of the respondents. Then the reader (you) read the précis of my interpretation, and will have your own interpretation of what occurred as marketing planning and learning were recorded in the logs. Therefore linguistic filters could warp reality. Respondents completed logs on a weekly basis, and therefore depended upon their memory to recall the events as they occurred during the previous week. Their records may contain mainly significant events, and leave out important details. They could also use the benefit of hindsight to adapt their interpretation of reality. In fact respondents were encouraged to reflect on why and how events occurred, and what actions should be taken to solve problems or achieve goals. Also as the interpreter of the reflective logs, the sequence in which the researcher reads them may also inform his view of the realities contained therein. So the researcher builds his own view as he reads each series of accounts, almost like reading a series of short stories – each with the same topic but written by different authors. The personality types of the respondents and that of the researcher may also taint any interpretations i.e. the way in which views are built and conclusions are drawn. Different respondents used different tones, and some logs began quite factual and became reflective as the respondent became au fait with the reflective approach to record keeping. In fact the mood of the respondents as they record their logs could distort the meaning within the text. At either a conscious or a sub-conscious level respondents may have their own agenda. Many comment on their staff and their managers, some containing more admiration than others. Respondents may be trying to shape an agenda when writing their entries and here again there may be an alteration in the way in which facts and reality are reported. There may also be a tendency on the part of the researcher to be swayed by more persuasive responses, and there may be barriers to clear thought hidden below the surface, not least the expectation on the part of the researcher that the project would be completed on time with plenty of interesting data to analyse. One of the most important filters in terms of learning log research is the filter of methodology. One could ask the questions at this point, do learning logs record the process of learning? Do a series of logs force the respondent to structure responses in such a way that they are artificially driven into a process of learning? Such questions are difficult to tackle, but it is accepted that either one view or the other may be closest to reality. Finally the filter of selection is in evidence. One must be selective in the use of evidence since not all data or text can be used in any final write up of learning log research. That which is not used, is in a way filtered out. Therefore there exists the filter of selection, since the researcher uses only those extracts that he feels go some way to support or counteract central themes.

4. Conclusions and recommendations

Learning logs can be used as a research method. They collect data from reflection, learning and experience. They are a suitable research method where research objectives or problems relate to personal experiences in the absence of the researcher. They are recommended to those investigating the experiences of individuals over a span of time, since logs are a longitudinal mode of research. Logs should be considered an interpretive method, and any empirical data emanating from logs must be treated with tremendous caution. The similarities between logs as an assessment and logs as research methods are clear, for obvious reasons. Both modes contain a similar structure, which is easily adapted from situation to situation, to suit individual students or individual research scenarios. They include initial audits that set the context, and SMART objectives are generated. A series of log entries contain the bulk of the data, embedded in personal reflection. Finally, the formative process is reflected upon and closed, with a summative final reflection. One can measure the difference, or gap, between the initial contextualisation and the final outcomes, with reflection, learning and experience taking

**Tim Friesner and Mike Hart**
place along the way. There will inevitably be a number of applications of learning logs as a method of research. However it is recognised here, that there is a potential for the approach to be understood in much more detail. Therefore the authors encourage potential future researchers to use the method and to share findings with the academic community. There are two clear areas for more research namely research into business and management disciplines, and research into higher education. Learning logs record reflection, experience and learning as it takes place in individuals, within organisations. So the fields of business and management could potentially employ logs as a research method. There is also a plentitude of rich and varied data available in the learning logs written by students for qualifications. It must be taken into account that permission should be obtained prior to any analysis, and that confidentiality is respected. Again researchers are encouraged to use interpretive techniques to gain insight into reflection, learning and experience. Any further research should also attempt to add to the interpretive techniques already advocated (Friesner and Hart 2005).

References

Franzosi, R (2004) From Words to Numbers: Narrative, Data and Social Science (Structural Analysis in the Social Sciences), Cambridge University Press
Ryan, M (2004) Narrative Across Media: The Languages of Storytelling, University of Nebraska Press
Interpretivism and the Pursuit of Research Legitimisation: An Integrated Approach to Single Case Design

Felicity Kelliher
School of Business, Waterford Institute of Technology, Ireland.
fkelliher@wit.ie

Abstract: While interpretive research is recognised for its value in providing contextual depth, results are often criticised in terms of validity, reliability and generalizability, referred to collectively as research legitimisation. This paper explores the criticisms levied on interpretive case studies and presents a research design that seeks to address these criticisms. The paper describes the research template developed by the author and applies it to a longitudinal case study carried out on a micro firm in the Republic of Ireland. Following some detailed evaluation and analysis the author concludes that legitimisation of an interpretative case study is improved when an integrative approach involving the combination of specific research techniques to relevant and appropriate standards is adopted.

Keywords: Interpretive case study, qualitative research, legitimisation

1. Introduction

Interpretivists believe that reality is not objectively determined, but is socially constructed (Husserl, 1965). The underlying assumption is that by placing people in their social contexts, there is greater opportunity to understand the perceptions they have of their own activities (Hussey & Hussey, 1997). By its nature, interpretivism promotes the value of qualitative data in pursuit of knowledge (Kaplan and Maxwell, 1994). In essence, this research paradigm is concerned with the uniqueness of a particular situation, contributing to the underlying pursuit of contextual depth (Myers, 1997). However, while interpretive research is recognised for its value in providing contextual depth, results are often criticised in terms of validity, reliability and the ability to generalise, referred to collectively as research legitimisation. These concerns are amplified in the single case scenario (Eisenhardt, 1989; Perry, 1998). In reality, all these issues are interdependent and reflect on the layered complexity of the phenomenon at hand:

- **Reliability** refers to the consistency or stability of a measure. Denzin (1970) states that multiple and independent methods should, if reaching the same conclusions, have greater reliability than a single methodological approach to a problem. This combination of methodologies in the study of the same phenomenon is known as triangulation. From an interpretive perspective, Eisenhardt (1989) recommends that the researcher start with a broad research question, establish systematic data collection and ensure case access to create strong triangulated measures. Qualitative research findings can be strengthened in this way by combining participant observation with interviews and documentary sources (Hammersley and Atkinson, 1983) in a single case.

- **Validity**: In terms of validation, qualitative research depends on the presentation of solid descriptive data, so that the researcher leads the reader to an understanding of the meaning of the experience under study (Stake, 1995). In essence, validation is an interpretive understanding of truth (Angen, 2000). Thus, triangulation is not a tool or a strategy of validation, but an alternative to validation in this context (Denzin & Lincoln, 2003). In a single case, data triangulation is particularly important in order to fortify validation in the absence of cross case comparison. Remenyi et al (1998) suggest using multiple data sources, establishing an identifiable chain of evidence, and having a draft reviewed by the key informants to strengthen construct validity in this regard.

- **Generalizability** refers to the extent to which the findings of the enquiry are more generally applicable outside the specifics of the situation studied (Robson, 2004). In qualitative terms, the research goal is to offer a case description (including data collection procedures) that would allow the reader to repeat the research process in another case (Kidder & Judd, 1986; Vaughan, 1992). Although a single case may not provide sufficient evidence to make robust generalisations, it can establish the existence of a phenomenon (Van Maanen, 1988), which is adequate for the purposes of exploratory research (Remenyi et al., 1998). Thus, a case can be generalizable to theoretical propositions (Yin, 1984), creating a distinction between
analytical and statistical generalizability (Yin, 2003).

The remainder of this paper is broken into four. Section one presents a research design that seeks to address the criticisms levied on the single interpretive case study. Section two describes the research template developed by the author and applies it to a longitudinal case study carried out on a micro firm in the Republic of Ireland. The concluding sections offer a perspective on interpretive data analysis and single case design evaluation. Following some detailed analysis, the author concludes that legitimisation of an interpretive case study is improved when an integrative approach involving the combination of specific research techniques to relevant and appropriate standards is adopted.

2. Case design

A research design is "an action plan for getting from here to there, where 'here' is the initial set of questions and 'there' are the set of answers" (Yin, 1994:19). In this study, the underlying research question sought to assess the learning impact of a critical incident on employees (both individually and collectively) in a micro enterprise. The longitudinal interpretive case was deemed the most appropriate method to facilitate a valid response to the proposed research question. For clarity, a case study can be described as the investigation of a contemporary phenomenon within a real-life context (Yin, 2003), while in-depth case studies are often the vehicle for interpretive investigations, where research involves frequent visits to the field site over an extended period of time (Walsham, 2002). Justification for the suitability of the chosen research instrument is founded on Hill and McGowan's (1999) work which suggests that small company research may be best done using a qualitative approach that includes participant observation, case studies, in-depth interviewing and the use of documentation. Considering learning is not a single event, but rather a phenomenon to be studied in past, present and future terms, observational evidence offered the most appropriate means of assessing the level of adjustment in this context (Sutton & Callahan, 1987). By applying this research method, causal assessment could be established through depth and time series analysis (Kidder, 1981; Kratchwill, 1978) rather than as a single point in time, offering the greatest potential for legitimisation of the research results.

Having established the primary research question, the resultant objectives and the research design ethos, the researcher prepared for data collection by developing a research protocol as recommended by Yin (2003). This research blueprint focused on what questions to study, what data were relevant, what data to collect, and how to analyse the results. It also encompasses the management criteria relating to the case, and the researcher’s role as the primary research instrument (Trauth, 2001). The protocol allows for a chain of evidence, ensuring increased reliability and reduced misperception at every stage of the research process:

Table 1: Research protocol

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research question</td>
<td>Assess the learning impact of a critical incident on employees (both individually and collectively) in a micro enterprise.</td>
</tr>
<tr>
<td>Research Method</td>
<td>A longitudinal case study</td>
</tr>
<tr>
<td>Critical Incident</td>
<td>Information System implementation in a micro firm</td>
</tr>
<tr>
<td>Case Selection Process</td>
<td>Environmental Criteria: Micro firm influenced by dominant supplier IS requirements and industry regulation. Internal Criteria: Micro firm, whose owner is in the pivotal managerial role as primary employee influencer and has an imminent/recent, IS implementation within the firm.</td>
</tr>
<tr>
<td>Case Access</td>
<td>Identify cases fulfilling the criteria in the research protocol. Negotiate full access to the case. Meet to establish researcher/employee rapport prior to the in-store IS implementation.</td>
</tr>
<tr>
<td>Research Instrument</td>
<td>Researcher as the primary research instrument in the application of research methods</td>
</tr>
<tr>
<td>Boundary device</td>
<td>Micro firm learning framework</td>
</tr>
<tr>
<td>Research Techniques</td>
<td>On-site observation and semi-structured interviews supported by reflective diaries independently generated by the case business owner and researcher, informal conversations, completion of learning questionnaires and focus groups, and the perusal of internal documentary evidence over a three year period.</td>
</tr>
<tr>
<td>Data Management</td>
<td>Audit trail of data, collection methods and process, including control of the research instrument’s influence on the studied environment, specifically the balance of observation/participatory action.</td>
</tr>
</tbody>
</table>

Adapted from Klein & Myers, 1999:80
Following an initial literature review, an IS implementation was selected as an appropriate critical incident in order to investigate its anticipated impact on employee learning, an impending outcome supported by Zuboff’s (1988:13) findings that “people who are working with technology for the first time are ripe with questions and insights regarding the distinct qualities of their experience”. As longitudinal case studies can offer a broader understanding of the ways in which people adapt technological systems for their own purposes, particularly over time, a three-year case duration was set in order to determine the causality links more explicitly in this context. Subsequently, the setting of case criteria offered the contextual landscape required to successfully test the research question and internal and external criteria were specified to assist in the identification of acceptable case candidates in this regard.

In order to successfully perform the case study, full and complete company access was vital. In this context, “random selection is neither necessary, nor even preferable” (Eisenhardt, 1989:537). Thus, following the identification of suitable cases, access was negotiated via personal contact with the owner. While negotiation took some time before access was granted, this process was invaluable in ensuring both parties were satisfied with the research terms of reference. Of particular importance was a mutual understanding of the amount of access being requested and the length of the study in elapsed time: three years is a significant investment, not only on the part of the researcher but also in terms of the case participants. In addition, a clear appreciation of the research objectives and the direct contribution in tangible and intangible terms that the research could make to the studied case should clarify boundaries relating to the research before it commences.

An overriding concern of the author’s was that the mere collection of in-depth case data does not provide theory concepts in and of themselves. This point is articulated by Zuboff (1988) who states that while observation could be considered first order constructions, researchers rely on good theory and insightful analysis for second order concepts in order to induce theory. Miles & Huberman’s (1994) suggest that conceptual frameworks can be used as boundary devices in this context, while Janesick (2002) recommends the development of working models and theories in action that explain the behaviour under study. Therefore, variables were identified early in the case and incorporated into a loose conceptual framework to focus the case investigation. Notably, Crossan et al. (1999) suggest that these tools can help define context and promote the move toward theory in the employee learning research milieu. Thus a framework’s latent legitimacy value is significant, particularly if the underlying case purpose is to provide a basis for theoretical replication.

3. Data collection template

A clear schedule of data collection activities was discussed with the micro firm owner at this point in the research process. The author also outlined a schedule of site visits, detailing each visit’s likely duration with the micro firm owner. A variety of data collection techniques allowed for a greater possibility of anomalies to be noted, and sought to accommodate limitations relating to individual techniques (Gallivan, 1997):

Table 2: Data collection activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
<th>Time line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation</td>
<td>Sporadic on-site observation over a three-year period</td>
<td>10/01 to 09/04</td>
</tr>
<tr>
<td>Literature review</td>
<td>Evolutionary process completed in tandem with the case study.</td>
<td>2002 to 2004</td>
</tr>
<tr>
<td>In-depth interviews/conversations</td>
<td>Owner interviews Shop floor manager interviews Employee conversations Employee focus groups</td>
<td>6/8 week intervals Quarterly 6/8 week intervals Annual</td>
</tr>
<tr>
<td>Diary maintenance</td>
<td>Researcher reflective diary Case owner reflective diary</td>
<td>10/01 to 09/04</td>
</tr>
<tr>
<td>Internal Document review</td>
<td>Complete review at the start of 2002, &amp; subsequent review at the end of 2002, 2003 and June 2004.</td>
<td>01/02 to 06/04</td>
</tr>
<tr>
<td>Industry review</td>
<td>Complete review in 2004</td>
<td>2004</td>
</tr>
</tbody>
</table>

- Observation: In order to comprehend the employee learning impact, it was deemed important to observe the micro firm and its employees over an extended period. One of the noticeable benefits of longitudinal observation was that general relations were often discovered in vivo, an advantage alluded to by Glaser & Strauss (1967). Thus, the longer the period spent in the field, the better (Gomm et al., 2002) in order to gain true insight into the workings of the micro enterprise in the context of the
research question. This point is particularly compelling in the single case scenario. It should be noted that longitudinal cases, even those that require periodic rather than continuous involvement are labour intensive (Mumford, 2001) and the basic ethos behind the interpretive epistemology necessitates the researcher’s direct involvement in all stages of the research process in order to allow the researcher the level of empathy required.

With reference to this case, the researcher was present for an initial two week period, encompassing the week directly preceding and the week directly following the IS implementation. The IS installation and upgrade process took eight to twelve hours daily, during which time the researcher was present for observation purposes, taking detailed field notes as required. Each ‘working’ day was followed by evenings documenting and interpreting the day’s observations in order to establish new lines of inquiry for the following day’s work. Subsequent two-day observation sessions at pre-defined six to eight-week intervals in year one and two, and at eight to ten-week intervals in year three allowed the researcher to document employee progress over a three year period providing substantial insight into organisational and social perspectives in the studied case over time. Setting the balance between participation and observation can be difficult. In this research, observer as participant (sometimes identified as interrupted involvement) was deemed the most beneficial. This decision was based on the fact that complete detachment would not create the necessary subject trust to realise the research objectives while complete participation could create conflict of interest. In terms of temporal interconnection, the goal was to establish learning in past, present and future terms. The underlying assumption is that learning causation is neither linear nor singular [founded on Pettigrew’s (1990) argument] but rather continual in the context of ongoing change.

Finally, observation invariably raises ethical dilemmas, particularly when conducted in a covert way. This dilemma was considered by the researcher prior to commencing the study and was discussed with the micro business owner at length. It was agreed that the researcher would be introduced, as a management consultant to the case employees, employed to establish the best means of optimising IS operation within the case organisation. The researcher’s particular interest in learning was explicit, having been raised at several staff meetings and documented in relevant meeting minutes throughout the observation period. Issues of privacy, confidentiality and the relationship between individual cost and scientific benefit were also discussed with both the business owner and the researcher’s peers in order to establish an appropriate balance in this regard. The primary tool used for dealing with risk was an informed consent form. The owner was provided with a simple written statement detailing the study’s research objectives, potential benefits and risks, rights to confidentiality, and termination procedures, in acknowledgement of relevant ethical aspects of the research process.

- **Literature Review:** A thorough literature review combined with initial case observations provided a number of potential dependent variables in the context of micro business IS adoption and the resultant impact on employee learning. These findings provided the basis for relevant environmental, organisational and individual employee variables when developing the learning framework.

- **Interviews:** Personal in-depth interviews were carried out with the case owner and Shop Floor Manager at pre-defined intervals throughout the case study. As the case sought to establish protocol, the interview schedule was revised as required based on participant feedback and researcher observation (as suggested by Glesne & Peshkin, 1992). These interviews were not recorded or directly transcribed as requested by the case participants, however, extensive notes were taken during the meeting and written up immediately (as recommended by Zuboff, 1988) to ensure optimum recall regarding the interview content. The goal was to ‘document carefully the practical contours of interaction in the varied circumstances in which they unfolded’ (Gubrium & Holstein, 2003: 229). It should be noted that recording the meaning of what is being said rather than the exact words of the respondent is more important in this context (Perry, 1998; Stake, 1995). These interviews were enhanced by periodic informal face-to-face conversations with each employee when the researcher was
on-site, and supplemented by annual group sessions conducted by the author. These interactions allowed employees to voice their individual opinions regarding the IS implementation and its impact on their work.

- **Diary Maintenance**: To achieve dialogical reasoning (Klein and Myers, 1999), researchers must confront the preconceptions that guided the original research design (Janesick, 2002) with the data that emerge through the research process. Underlying philosophical assumptions should be transparent in this regard. Specifically, that the researcher is predisposed to factors of perception and prior theories, as well as prior expert knowledge (Yin, 1994), all of which influence what we take to be factual observations. Geertz (1973: 9) articulated this issue as thus: "what we call data is really our own constructions of other people's constructions of what they and their compatriots are up to". To help overcome these issues, Trauth (1997: 241) recommends, “Self-conscious questioning of [the researchers] own assumptions [to] bring into consciousness the emotional and intellectual reactions to experiences and observations”. In the case study, diaries were used as a supplementary research tool valuable in the pursuit of reliable research by providing an essential audit trail in the research context. Several writers advocate the maintenance of a reflective diary as a rigorous documentary tool in this context (for example: Glaser & Strauss, 1967; Janesick, 2002; Stake, 1995). Thus, the researcher maintained a reflective diary separate to the non-reflective recording of observations throughout the case study. This amounted to a personal journal of the research process, specifically the recording of emergent ideas and results, reflections on personal and case participant learning, and an ongoing examination of personal attitude that proved invaluable when analysing the case data. Separately, the micro firm owner maintained an IS related diary throughout the case duration. Here, the diary provided a simple record of events and sought to assist in the identification of certain activities and the frequency of occurrence. Unexpectedly, the owner diary offered new perspectives that were not articulated in either the interview or the observation process.

- **Internal document review**: Having agreed access with the owner, the author carried out a thorough review of the case site's internal documentation including internal quality manuals, IS training documentation and user manuals, 2002 to 2004 business diaries, IS vendor correspondence, wholesale supplier literature, and a review of all historic staff meeting file notes and notes included by the owner in the staff's monthly pay packet. Each record provided valuable data relating to the culture of the organisation and its relationship with the IS vendor and dominant industry suppliers. They also offered supplementary knowledge relating to employee learning focus over time.

- **Industry review**: Separate to the case site data collection, the researcher interviewed an industry expert to establish the case's external environment in terms of political, legal and industry pressures placed on a micro firm operating within the sector. Three in-depth interviews were carried out at the start of the case, each lasting approximately two hours, in order to establish relevant criteria relating to the environmental aspect of the learning framework. A thorough review of all relevant public documentation relating to the sector was also completed in this context. Documentary evidence sought to 'corroborate and augment evidence from other sources', (Yin, 2003:81). In essence, documentary evidence provides context, particularly in a complex environment, creating greater validity and reliability of the research and the resultant framework.

4. **Data analysis and interpretation**

The philosophical links remain at the data analysis stage of the research process. Considering this research's interpretive stance, the ultimate goal is to describe the context in which events occur. The analytical goal is to make sense of the whole [situation] and the relationship between people, the organisation and technology (Myers & Avison, 2002). Therefore, the underlying philosophy dictates an iterative process of data collection and data analysis (Eisenhardt, 1989; Walsham, 2002), which are tested and modified through cycles of additional data collection and analysis until an adequately coherent interpretation is reached (Glaser & Strauss, 1967). Consequently, the researcher sought to overcome the temptation to convert qualitative data into numbers once it had been collected, in order to preserve the richness of the data and give a holistic view of the research context. While interpretive research does not subscribe to the idea that a pre-
d) In step four, the researcher abstracts the
essences from the text. Essences are
wholly subjective gestalts of what is
learned from studying the phenomenon,
and requires creativity, intuition and
reflection. Thus, analysis depends on an
investigator’s own style of rigorous
thinking (Yin, 2003). The pre-mentioned
reflective diary proved invaluable in this
context, not only in terms of the resultant
content, but also in the process of
reflective thought required to complete the
diary task. Finally, evidence is presented
in a narrative form, supported by evidence
from the statements and behaviours
recorded in field notes, diaries and
interviews (a process supported by
Janesick, 2002; and Kidder & Judd, 1986). Internal validity is established where
several pieces of information from the
same case point to a theoretical
proposition (Stake, 1995) or empirical
assertion (Janesick, 2002). Therefore,
insights are validated with rich
descriptions, direct quotes from
participants, and practitioner review of the
interpretation (Lacity & Hirschheim, 1993)
in order to build the thick description
(Geertz, 1973) sought under this analytical
ethos.

With reference to the research report, it can be
linear-analytic, comparative, chronological or
theory-inducing (Yin, 1994) dependent on the
research philosophy and underlying
assumptions adopted. In this instance,
chronological reporting provided the greatest
insight into the evolving impact of the critical
incident on learning over the case duration.
Specifically, a chronological record over a
prolonged period should offer greater insight
into learning as an evolving phenomenon. As
this action is founded in the intentional
research process detailed above, the
researcher planned for the final report creation
from the start of the investigation (as advised
by Yin, 2003). This iterative approach to data
collection and analysis created an important
cycle of discovery within the case. Finally, it
was important to establish a strategy whereby
the micro firm could gain from the research
without impeding on the research reliability
and validity from an academic standpoint.
Academic rigour had to be acknowledged in
this context. Thus, once documented, the
micro firm owner reviewed a draft copy of the
case study to ensure the document contents
were complete and accurate from her
perspective. It was also agreed that the author
would perform an IS audit as part of the
observation process and produce a feedback
report on the impact of the learning plans and
training programmes on each employee from
an individual and collective perspective in this
context.

5. Design evaluation

The chain of evidence provided by the
established research protocol, the subsequent
research template and the integrated
application of appropriate research techniques
discussed in the preceding sections of this
paper sought to legitimise the research results
throughout the duration of the case.
As previously stated, complete case access was paramount and its importance cannot be over-emphasised in the context of this paper. The researcher’s presence at the IS implementation and subsequent upgrades proved vital as it offered contextual knowledge relating to employee learning. Specifically, the researcher observed and documented the individual employees’ initial reaction to the information system during the installation and directly following its implementation. This initial investigation sought to identify the relationships and interdependencies between individual learning and collective learning within the micro enterprise, in the context of the external environmental influences and how these elements shape such relationships; culminating in the creation of the learning framework. Having established that the IS implementation required significant adjustment in the employee’s knowledge, observational evidence, combined with internal interviews, focus groups and informal conversations, offered the most appropriate means of assessing the level of employee adjustment in context. The value of multiple perspectives over the initial observation period established broad themes and provided for greater data reliability when analysing the case results.

Building staff/researcher empathy proved valuable in terms of mutual comfort with individual employees during the study. The ambition here was to get as close as possible to the world of the micro firm’s decision catalyst and IS implementation and to interpret this world and its problems from the inside in order to describe both the unique and typical experiences and events within this environment as bases for theory (as argued by Dalton, 1959). Observed data was documented throughout the case study and periodically checked by participants and research colleagues to reinforce objective reporting. The goal was rigorous and systematic data collection, without excluding serendipitous information (Kaplan & Maxwell, 1994).

The complementary data collection techniques used throughout the case proved particularly valuable in the pursuit of research legitimisation. Reviewing the firm’s internal documentation provided greater insight into the firm’s internal culture. It also afforded an appreciation of the micro firm’s ongoing relationship with both the IS vendor and dominant industry suppliers, while the industry expert interviews gave a holistic view of the micro firm’s external business environment.
Maintaining reflective diaries as a complimentary research tool throughout the research process offered an additional audit trail in relation to theme development. Specifically, the researcher’s reflective diary acknowledged and identified personal reference frames on a continuous basis while the owner’s diary offered individual perspective without the imposition of the researcher’s own reference frame during data collection. Lastly, informal conversations with individual staff members and focus group discussions sought to identify individual employee’s interpretation of the IS implementation and subsequent learning needs, rather than being restricted to those of the business owner alone. In this context, the principle of multiple interpretations is of heuristic value because it leads to probing beneath the surface; a benefit alluded to by Hussey & Hussey (1997) and Kaplan & Maxwell (1994) in the introduction.

6. Conclusion

This case study sought to establish a typical case supported by a detailed research protocol (Yin, 1994). The underlying purpose of the case design was to provide a basis for theoretical replication alluded to by Yin (1984). The multiple research techniques applied in this case offered reliability via triangulated results. Finally, the pre-established data collection template provided for a standard approach to data collection and analysis in pursuit of research validation. As a note of caution, this form of research is not just a matter of observing the subject matter and analysing the results at a distance at some later date. It involves observing, participating, talking, checking, understanding and making interpretations over an extended timeframe, all of which are required if the observer is to share and understand important parts of the employee’s experience. In addition, case participants can be seen as interpreters and analysts in their own right, whose horizon is changed by the researcher’s interaction with them (Klein and Myers, 1999). It is inevitable that as the researcher establishes mutual comfort over time, and that they interact with the case subjects at a social as well as professional level. In consequence, reporting on the case participants impartially can prove to be a difficult task. Thus, pre-established standards in each aspect of the case design, protocol, data collection, analysis and interpretation provide for greater legitimisation of research outcomes in this regard.

References

Angen, M.J., 2000, Evaluating interpretive enquiry: reviewing the validity debate and opening the dialogue, Qualitative Health Research, 10(3), p378-95
Dalton, M., 1959, Men who manage, NY: Wiley
Janesick, V.J., 2002, The choreography of qualitative research design, in: Denzin,


Miles, M.B. and Huberman, A.M., 1994, Qualitative Data Analysis – an expanded sourcebook, CA: SAGE.


Tell me a Story –A way to Knowledge

Dan Remenyi
School of Systems and Data Studies, Trinity College Dublin, Ireland
dan.remenyi@tcd.ie

Abstract: A narrative or a story (and these terms are synonyms) is a fundamental way of understanding our environment and relationships in it and thus it is a key feature of sound research whatever methodological approach has been taken. The skills of story telling are important in both qualitative or interpretive and quantitative research. But it comes into its own in qualitative research where story telling especially as it appears in case studies and action research is of prime importance. In quantitative research a story or narrative is also required to contextualize the work. The argument developed in this paper balances the idea that numerical analysis underpins the most powerful research paradigms.

Keywords and phrases: Story telling, narrative, qualitative research, quantitative research, rhetoric, argument

Tell me a story, tell me a story
Tell me a story before I go to bed;
You promised me, you said you would,
You gotta give in 'cause I've been good,
Tell me a story before I go to bed.
http://www.poetrylibrary.org.uk/poetry/quote/txtreply.jsp?quote_id=2228

1. Introduction

This is a speculative paper which consists of some reflection on the nature of knowledge created by qualitative research in the business and management field of study and discusses the critical importance of story telling in the research process. The paper is written with the intention of stimulating discuss concerning the importance of the narrative in research. The argument presented here is developed to balance the idea that numerical analysis underpins the most powerful research paradigms. It is not the intention of this paper to attempt to denigrate the importance of quantitative research. In fact numerical analysis is sometimes said to be required if we are to have ‘true’ scientific knowledge which is regarded as the highest form of knowing (Mays and Pope 1995). Numerical based research, sometimes referred to as the positivistic approach to research plays a very important role, if not a dominant role, in business and management studies research. This is more due to historical reasons than the inherent strength of this approach and this situation is changing. The dominance of numbers in research in general is well expressed in the words of Gould (1992):-

The mystique of science proclaims that numbers are the ultimate test of objectivity.

2. Quantitative or numerical research

But seeing numbers as the ultimate form of objectivity does not reflect the personal values of Gould concerning the appropriateness of the authority of quantitative analysis. Making his position on quantitative research clear Gould (1997) alerts his readers to the potential difficulties which can arise in the research process when he says:-

Our searches for numerical order lead as often to terminal nuttiness as to profound insight.

The need to find numerical order in the world around us is an old one and can be traced back to Pythagoras (Koestler 1959). In fact Koestler points out that to the Greeks the ‘mathematisation’ of experience meant enrichment of knowledge and certainly as far as the problems tackled by the early Greeks was concerned, this was mostly the case. There is little doubt that the technology, at least in the traditional sense of the word technology, on which our modern world is built, has been honed through ‘mathematisation’. Needham (1988) goes so far as to say

'Modern [as opposed to mediaeval or ancient] science is the mathematization of hypotheses about nature ....... combined with rigorous experimentation'

This view would be supported by the majority of the scientific community, especially those in
the physical and life sciences\textsuperscript{1}. But clearly this is not the whole story as numbers themselves have no intrinsic value. In fact numbers can be really misleading. And this is pointed out by Paulos (1998) as he discusses how some scholars try to hard to find numerical or logical trends or patterns in data when he describes the results of equidistant letter sequences (ELS\textsuperscript{2}) research. Paulos cautions us against looking too hard to find patterns.

Despite its critics, mathematics and its companion statistics\textsuperscript{3} remain a very powerful tool set for the researcher. They have enriched our research and can lead to highly useful results. Methods of research that do not use mathematical or statistical techniques are sometimes seen as being soft or inferior, i.e. not being that powerful or reliable or even valid, and this can lead to heated or even acrimonious debate (Webber 2004). But on the other hand the mathematical and statistical approach to research clearly has it limitations which are unfortunately not always that obvious.

In fact there is a delicate relationship between quantitative research and other forms of research, of which qualitative research is probably the most popular and this relationship needs exploring. Our ability to understand the world and thus create knowledge was originally through verbal expression and this pre-dates our numerical facility by many millennia. Much of our knowledge has been created through non-numerical research strategies. It is interesting to note that it was only in the 15\textsuperscript{th} century that modern mathematical notation came into use (Ball 2001). But even then many of the early scientists and researchers were predominantly descriptive. As science and research progressed so there was a need to tackle increasingly more complex issues which required expression in specialised language, some times referred to as the ‘algebra’ of logic and here mathematics played a key role. Exactly when statistics began to play a dominant role in the research of the physical, life and behavioural sciences is not clear. There was certainly much activity in the second half of the 19\textsuperscript{th} century by individuals such as Galton, Edgeworth, Pearson and Yule. But Salburg (2001) declares that:-

I would prefer to date the statistical revolution to the work of Karl Pearson\textsuperscript{4} in the 1890s.

Despite the strides made with numerical analysis it needs to be remembered that a material amount of what can be done with mathematics and statistics in research can also be done with verbal descriptions, explanations and argument (Paulos 1999).

As a language, mathematics communicates abstract and complex ideas in order to reduce the ambiguity which is inherent in natural languages\textsuperscript{5} and to provide us with an ability to more easily discern patterns among variables. For this reason it is often described as the language of science. Through its precision and economy of expression mathematics facilitates the definition and manipulation of abstract and complex ideas, which could not be easily accommodated in natural languages. Mathematics uses conventions to express relationships in symbols rather than words according to a strict code of logic. As alluded to above, in the physical and life sciences the power of mathematics to help solve problems and understand our environment is unquestionable. But as Paulos (1999) points out the code of logic on which mathematics and statistics is based is available to the qualitative researcher and although not using mathematical and statistical techniques qualitative research can derive sound conclusions through using similar logical procedures.

3. Complexity in management and business studies

However, considering research in the world of business and management studies the situation offers a number of special challenges. Although it is not true in all cases the complexity of much of the business and management world makes its formal description using mathematical tools such as

---

\textsuperscript{1} There would be considerable less agreement among social scientists, although there would be some who would argue for this type of research.

\textsuperscript{2} ELS are used to find hidden meanings in texts especially biblical texts. See http://cs.anu.edu.au/~bdm/dilugim/Nations/WRB2/ Some would argue that ELS try to push us beyond any reasonable interpretation of our cognitive ability.

\textsuperscript{3} Although there is no universal agreement on this point, statistics is sometimes described as applied mathematics.

\textsuperscript{4} Karl Pearson (1857-1936) spent most of his life as Professor of Applied Mathematics at University College London and famous for his Pearson Product-Moment Correlation Coefficient.

\textsuperscript{5} As well as being seen as a language, mathematics may be regarded as an intellectual universe in its own right. In this respect it goes beyond facilitating the communication and manipulation of ideas as it brings forth concepts for which there is no easily describable linguistic meaning. Concepts such as these often provide useful insights into the natural world as well as the social world as well as artifacts.
symbolic notation or even standard logical rules quite difficult. When such methods are used we often have to simplify what we are doing or what we are working with. Although simple models can be useful, sometimes this is done to a point where a significant amount of the richness of the situation is lost and when that happens their utility has been diluted. When this occurs researchers or scholars tend to argue that the main point of the mathematical representation or model is to communicate the general ideas which have been symbolically expressed in it rather than using the model for further analysis or optimization.

To attempt to describe or understand how an organisation’s promotion policy actually works is often difficult to comprehend when reduced to a number of rules or even guidelines. Even describing something like the relationship between various levels of output and costs can be quite challenging in symbolic notation. When we do this we have to be even more careful than usual to define the variables closely and to explicitly state the assumptions underpinning them and to clarify the relationships between them. There is also the challenging issue of across what range of values these relationships are expected to be valid or robust. Of course variables also have to be defined in qualitative research but perhaps with some lesser degree of detail.

3.1 The narrative

The complexity of some situations can be more comprehensively described, understood and communicated using metaphors, similes and allegories and other verbal expressions. In short a story or a narrative (or indeed a picture) can sometimes deliver much better and clearer message than any other form of communication. This is of course how our pre-literate ancestors remembered who they were and what they had done and what they knew. Even today in some societies the story is a celebrated vehicle of retaining and expressing its sense of itself or its cultural.

In order to minimize any confusion it is necessary to point out that in this context the term story and narrative are used interchangeably. According to the dictionary included in the CD version of the Encyclopedia Britannica 2005, a synonym of story is ‘anecdote, narration, narrative, tale’.

In terms of our current management culture, story telling has been taken up by Denning (2000) and others as a specific management tool and as a central platform of knowledge management. It is being rediscovered in this new context where it is being proposed as a corporate management stratagem or tool.

3.2 The story in business

For an example of the power of the use of a story in the business and management setting consider the following as an account of how entrepreneurship and perseverance may be described and understood by reference to a narrative about Chester F. Carlson and the dry paper copier.

“Chester F. Carlson having invented the dry paper copier in 1939 offered the rights to the process to every important office-equipment company in the USA, of which there were quite a number. Despite the clear advantage this new process had over the old wet process, he was turned down by every one of them. However in 1947 he managed to sell the rights of his revolutionary process to a small firm called the Haloid Company that made photographic paper. Unfortunately they did not have the resources to develop the dry paper copier on their own – at that time. It then took ten years for them to improve the product with Carlson having to canvass door to door in Rochester to sell shares to raise the money he required to complete the development of his concept. He also sold shares to employees.

6 This is not to say that there will not be some value in describing some business relationships in mathematical notation and even constructing computer models of these relationships.

7 According to Hawkins (1993), “I was sure that nearly everyone was interested in how the universe operates, but most people cannot follow mathematical equations – I don’t much care for equations myself. This is partly because it is difficult for me to write them down but mainly because I don’t have an intuitive feeling for equations. Instead, I think in pictorial terms, and my aim in the book was to describe these mental images in words, with the help of familiar analogies and a few diagrams.”

8 We can only speculate when stories were first told but cave drawings have been dated to around 30,000 years ago. These are regarded as a pre-literate record of story telling.

The Chester F. Carlson is told on many websites one of which is http://news.bbc.co.uk/1/hi/sci/tech/2012385.stm

The Chester F. Carlson is told on many websites one of which is http://www.yesterdaysoffice.com/index.cfm?fuseaction=Sh owArticle&articleid=26

ISSN 1477-7029
Eventually in 1959 the Haloid Company whose name was by then changed to Xerox (which is the Greek word for dry) developed the 914 plain paper copier. The 914 was an instant success, and has been called the most successful commercial product in history. Xerox was hailed the greatest big-business success of the 1960s. Many of the individuals who had bought shares from Carlson become very wealthy and were known as the Rochester Millionaires.”

If you now make a list of the business learning points from this narrative it will become apparent just how little direct impact such a representation of the above event has relative to telling the story itself. Then if you wish you take the business learning points and re-state them using formal symbols in the form of a mathematical equation. You can indeed create some meaning for some people this way but it is clearly limited. This story, especially when told with a little bit of feeling and enthusiasm, is much more powerful than any formal treaties on the virtues of entrepreneurship and perseverance. The story telling approach seems to be a fundamental human way of sharing knowledge. This is so deeply rooted in human nature that it has been suggested it may be part of our genetic code. Denning (2001) said that dogs sniff: humans tell stories. Gould (1997) expresses a similar idea when he said:

Humans are story telling creatures pre-eminently. We organise the world as a set of tales.

4. Other non-mathematical communications

Other examples of the power of non-mathematical communications may be seen in the visual arts. As mentioned above the visual representations have been used since earliest times in the form of cave drawings, which are believed to substantially pre-date any form of literacy. Diamond (1998) considers writing to be a phenomenon developed in the past 10,000 years whereas cave drawings are considerably older.

Of course, the visual arts are a combination of images which are in a sense symbols, but they are generally not abstract symbols. Take for example Vincent Van Gogh’s painting created in Saint-Rémy in June, 1889 and called Starry Night.


It is unlikely that any form of words or abstract symbols could be as effective in describing how Van Gogh saw a cloudless star lit night as well as the oils on this canvas. This extraordinary painting captures the painter’s feelings about and his mood concerning the wonder of the heavens on a completely clear night in a totally unique and completely convincing way. It is obvious that the ideas expressed in this painting are not easily expressed in any other format or way. If any proof is needed then try to write a few sentences or even a long treatise to describe this painting. It is extremely difficult, certainly for the majority of people, to match in word form the message communicated by the artist – never mind trying to reduce this painting to some sort of formal mathematical notation.

5. The narrative and the research cycle

To return to the role of the narrative in science and research it is useful to note that research projects invariably begin with a discussion of a problem. A dialogue or a discourse often triggers the researcher’s interest in a field of study or a topic to be researched. From this

\footnote{Clearly there are exceptions to this, but the majority of art would be thought of as being composed of images which were different to the abstract symbols used in mathematics.}

\footnote{It is interesting to speculate if Shakespeare could have written, say, a sonnet which could have had a similar impact as this painting.}
discussion a research question is found which is invariably expressed verbally.\footnote{Of course there may be exceptions to this as in pure mathematics.}

At the other end of the research cycle, no matter how mathematically based the work has been, when the results have been obtained and the findings are being reviewed or interpreted, we are once again heavily involved in discussion and a narrative takes over again. When the findings are expressed in numbers or in abstract symbols they generally have little meaning, except to the person who has developed them. The need for the narrative is well explained by Paulos (1999) who points out:

> Without an ambient story, background knowledge, and some indication of the providence of the statistics, it is impossible to evaluate their validity. Common sense and informal logic are as essential to the task as an understanding of the formal statistical notions.

Thus quantitative research is topped and tailed by narrative or a story. Looking at this situation diagrammatically in Figure 1 one can see that quantitative or numerical research represents only a relatively small part of the total research effort. Quantitative research is a sub-set of research in general and needs to be clearly understood as such. Thus it has no privileged claim to be a better form of research. Numbers do not represent a ‘truer’ or richer form of scientific knowledge.

![Figure 1: Numerical research as a subset of research in general](image)

6. Argument or rhetoric

When it comes to research findings the most important issue is perhaps the strength of the argument the researcher poses to support his or her work. An older word for argument is rhetoric and this term is again gaining popularity in the field of business and management research. Rhetoric is defined by Sarton (1980) as:

> the art of expressive and persuasive speech.

Although the word speech is used in Sarton’s definition, it is likely in today’s research environment to be more important that the researcher writes expressively and persuasively. Even the highest quality research without an appropriately powerful rhetoric is unlikely to make a great impact on society. As Walsham (2002) pointed out:

> Van Maanen suggests that the researcher must try to persuade by ‘presenting a coherent point of view told with grace, wit and felicity’.

It is easy to point out the importance of Rhetoric but it is not necessarily easy to define quality performance in this regard. Many researchers find it quite difficult to craft an expressive and persuasive argument. Many researchers would be completely lost when it came to trying to create a piece of text which possessed the qualities of ‘grace, wit and felicity’.

Not everyone can be taught to write well and to write with a high degree of conviction and persuasiveness is probably a gift which not many people possess. Nonetheless there is no reason why researchers should not aim for this.

7. Different styles of story telling

There are of course a number of different styles of story telling and researchers may find that their ability to develop a sound argument is stronger in one style. Van der Blonk (2004) defines four basic styles of case study writing which he calls Chronology, Play, Biography and Voices. These styles are derived from two sets of axes which he proposes. Van der Blonk suggests that a case study may be written as a monologue or a ‘multilogue’. Also may be written in considerable detail demonstrating the complexity of the situation\footnote{This full account of a situation or an interview is sometimes referred to as a primary narrative.} or it may be reduced in volume showing only a summary of the events and discussions\footnote{This summarised or reduced account is sometimes referred to as a secondary narrative.}.

If the case study is written as a monologue and the complexity is retained then Van der
Blonk refers to this as a Biography. If the case study is written as a monologue and the complexity is reduced then he refers to this as a Chronology. This could also be referred to as a History or Historiography. On the other hand if the case study is written as a ‘multilogue’ and the complexity is reduced then he refers to this as a Play in the sense of a theatrical event. If the case study is written as a ‘multilogue’ and the complexity is kept then he refers to this as Voices. These different approaches are shown in Figure 2.

**Figure 2:** The Van der Blonk taxonomy of case study styles

8. Summary and conclusions

In summary symbolic forms of communications such as mathematics or statistics are powerful, but limited, even in the physical and life sciences. They are especially potent in describing the rather tightly defined variables and relationships in the physical world. This approach is very powerful when it comes to manipulating data and when it is necessary to find trends or relationships among variables especially for the purposes of prediction. But in the end the findings of this type of research have to be interpreted and discussed and this is more effectively achieved by a narrative or story. When it comes to exploring the social sciences or business and management studies, these symbolic forms of communications are often inadequate. The use of mathematical or statistical approaches was commented on by Jung (1995) when he said:-

*Science works with concepts of averages which are far too general to do justice to the subjective variety of an individual life.*

However it needs to be admitted that this reflection of Jung’s, valid as it may be, does not really forward the purposes of the social scientist. It would be counter productive for us to say that all individuals and organizations were so various and our understanding of them so subjective that no generalizations were possible. What is necessary is to find a middle way between the ‘concepts of averages’ and the ‘soft philosophical black hole’ at the other end of the presumed spectrum.

Perhaps at the end of the day the primary issue is what techniques help us to better understand the world about us. In his book on quantum mechanics, Malin (2001) points out:-

*I discovered that a conversational mode would be helpful. Therefore I introduced two fictional characters.*

If this is what is required to create and disseminate knowledge then it is an acceptable method (Feyerabend 1990).

Of course this is not to say that we do not need abstract or mathematical and theoretical explanation in all forms of social science including business and management studies. But it is certainly important to support our abstract and theoretical thinking with powerful illustrative descriptions which are best delivered through stories.

It is for this reason that the story has a major role to play in the understanding the processes and findings of research. As narratives are cornerstones of understanding, it may therefore be said that stories themselves create knowledge.

Looking at this carefully we can see that the story is used not only to describe a phenomenon, but also to place it in its context and to explain the relevant adjacent or interconnecting issues and relationships. Perhaps as Malin (2001) implies the greater the complexity being dealt with the greater the need to use a story. In general the closer the story relates to actual events and their repercussions the more authentic it is and thus the more likely it is that it will have direct theoretical implications. In the extreme the story is the theory. This view is supported by Sutton and Staw, (1995) when they said

*Theory is about the connections among phenomena, a story why acts, events, structure, and thoughts occur. Theory emphasis the nature of causal relationships, identifying what comes first as well as the timing of such events.*

A good story travels well – its message resounds with many people (Denning 2000). It is enduring and thus it lasts a long time. And
furthermore it does not really need detailed numerical analysis.

Finally to make sense of the epigram at the start of this paper, story telling transcends many social divides including age and can actually be most enjoyable for both the teller and the listener especially when it is recounted with grace, wit and felicity.

References

Denning S, 2000, The Springboard, Butterworth Heinemann, Boston
Denning S, 2001, comment made during a presentation at Henley Management College.
Malin S, 2001, Nature Loves to Hide, OUP, NY
Mays N and C Pope, Qualitative Research: Rigour and qualitative research, BMJ 1995;311:109-112 (8 July)
van der Blonk H, 2002, Writing Case Studies In Information Systems Research, European Conference on Research Methods, Reading University, UK
Webber R, 2004, Editor’s Comments, Vol. 28, No 1, iii-xiii, MISQ.
Applying a Behavioural Simulation for the Collection of Data

Kristina Risom Jespersen
The Aarhus School of Business, Department of Marketing and Statistics, Denmark
krr@asb.dk http://www.asb.dk/staff/mis/krr.aspx

Abstract: To collect real-time data as opposed to retrospective data requires new methodological traits. One possibility is the use of behavioral simulations that synthesize the self-administered questionnaire, experimental designs, role-playing and scenarios. Supported by Web technology this new data collection methodology proves itself valid and with high appeal to respondents.

Keywords: Real-time data collection, simulation, Web technology

1. Introduction

As the complexity of problems and the speed of changes increase for companies, the accuracy of research relies on the measurement of behavior close to the real world of the respondents. Recent reviews in various research fields of the social sciences have pointed at the need for more interactive data collecting methods (Zaltman 1997; Englis and Solomon 2000). The collection of real-time data as opposed to retrospective data requires new methodological traits.

Another aspect stressing the development of new data collection methodologies is the prediction that response rates of e-mail surveys will follow that of paper surveys (Shenan 2001). To obtain data, to make people invest time in research projects, research has to follow the information ‘rules’ of society. Here the increased number of media, the habitually use of the Internet and mobile phones as well as a TV set on multiple channels, makes heavy calls on research to be an interesting experience for researchers to be able to collect the important and necessary data. Not because people do not find science important but because time is a scarce resource in heavy demand. Hence, investing time in research projects has to give the respondents something more than being polite and helpful towards the researcher.

The purpose of this paper is to discuss and demonstrate the benefits and challenges of behavioral simulations supported by Web technology as a new methodological trait. Specifically how behavioral simulations as data collecting vehicles also act as learning tools for the respondents thereby giving immediate return in the time invested.

The concerns of this paper are the methodological contribution of simulations, the benefits of Web technology followed by a discussion of the validity of data collected with simulations. Finally, a designed behavioral simulation is presented as an example, before stating implications and future trends.

2. The contribution of a behavioural simulation

To understand the methodological contribution of a simulation it is beneficial to look at the research area of agent-based simulations. The first agent-based computational economic models were influenced by experimental economics, which at the time mainly focused on markets and games composed of real human agents (Chen 2004). A criticism of experimental economics on information utility states that the problem lies with the utility functions and whether these resemble the true utilities of information, because human agents are subject to biases, errors, and misconceptions (Einhorn and Hogarth 1981; Feldman and March 1981; Chaturvedi, Mehta et al. 2004). Parallel to this, behavioral economic research proved that the biases creating inference in experimental economics are similar to those in real life. Humans optimize their behavior as assumed in experimental economics, but from a satisfying criterion because humans are bounded rational (Newell and Simon 1972; Bettman, Payne et al. 1993; March and Simon 1994).

For the understanding of the effect of bounded rationality in economic models, the research area of agent-based simulations emerged (Chen 2004) as a way to observe actual behavior as opposed to behavior deduced from a set of axioms. It is this objective to observe actual behavior that makes simulations a useful platform for the need to collect real-time data in social sciences. The spin-off from behavioral simulations is that participants are able to learn from them; thereby the models exhibit intelligence.
Adapting the knowledge on agent-based modeling (Chaturvedi, Mehta et al. 2004; Chen 2004), a behavioral simulation synergizes the benefits of self-administered questionnaires and the experimental design, and furthermore introduces role-playing (Dabholkar 1994; Armstrong 2000) and scenario (Frederickson 1984; Frederickson 1985; Eroglu 1987; Sanderson and Sanderson 2000; White, Varadarajan et al. 2002) strategies as very effective methods to ensure high interaction with the respondents.

The aim of the questionnaire is two-fold in ensuring that information about the decision behavior in companies is retained and in validating the observed behavior of respondents in the simulation as to whether it resembles behavior in companies. Hence, with the questionnaire the specification of the decision parameters is ensured.

Where a questionnaire represents the scope, an experiment focuses on depth by offering the opportunity to give a standardized and controlled presentation of the surroundings (the environment) of the decision situation (Perkins and Rao 1990). Hence, contrary to the questionnaire the experiment introduces activity and dynamics though the setting is more simplified than the ‘real world’; for example, the technological development, the competitive environment, or the uncertainty of demand. As such exogenous factors to the decision-making process can be specified and controlled in the simulation through experimental design. Using the questionnaire for measurement and the experiment for variation, these two methodologies provide the environment for the simulation. To make the simulation interactive role-playing and scenarios are essential methods.

Traditionally, role-playing has been used to forecast decisions in conflict situations among interacting groups, but it can also be used to predict decisions by an individual not interacting directly with others (Armstrong 2000). More importantly the similarity between laboratory research and role-playing is well documented (Dabholkar 1994). The advantage of role-playing is that roles influence a person’s perception of a situation. Participants are asked to engage in the role description and then either to imagine their actions or to act them out - in both cases as they would in fact do, i.e. managers should not play customers or vice-versa. The key is to make the role realistic (Eroglu 1987; Armstrong 2000). Therefore, with role-play the decision-maker in the simulation can be specified.

Table 1 provides some basic design principles.

<table>
<thead>
<tr>
<th>Table 1: Basic design principles for role-playing* (Jespersen 2004; Armstrong 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The role-players should be similar to those being represented. Meaning that a role-player should act as him/herself.</td>
</tr>
<tr>
<td>Role-players should read instructions for their roles before reading about the situation.</td>
</tr>
<tr>
<td>The administrator: provides short yet comprehensive descriptions and creates realistic surroundings in order to provide a realistic enactment of the situation.</td>
</tr>
</tbody>
</table>

The scenario is the situation in which the role-play is acted out by the participants and is therefore the heart of the simulation. The strength of using scenarios as frames for the decision-making process is that it makes the respondents relate more directly to the posed subjects, and to a high degree this results in more accurate responses (Eroglu 1987). The ability of a scenario is threefold in that it (i) increases interest in participation, (ii) makes it possible to create a realistic context, and (iii) provides all respondents with a standard stimulus (Frederickson 1984; Frederickson 1985; White, Varadarajan et al. 2002). Also when respondents are presented with a scenario before decision-making, their attention is guided to the relevant problem area. From this viewpoint, the scenario can successfully complement the experimental design by providing control of the decision-making process investigated. Thus, the scenario defines the behavior of the artificial reality in the model.

The methodological contribution of simulations is a symbiosis of four acknowledged research methods. At the same time, these methods make the simulation very complex. Thus the key to make a simulation suitable for data collection purposes is Web technology, because Web technology can carry out the large database structure upholding the simulation and making the simulation active and dynamic. Therefore, the possibilities with a simulation as a platform for a data collection methodology are closely linked to the use of Web technology when collecting data.
3. The benefits of Web technology

The use of computers for the collection of data through various methods is well known. As early as in the 1970s the first computerized experiment was launched. Especially within the field of decision-making, experiments have been computer-interactive from early on (Connolly and Thorn 1987). In a study comparing Web-versions with laboratory studies, Krantz and Dalal (2000) found a surprising match for surveys, scales and experimental variables thereby stressing the power of the Web as both a research medium and as experimental medium. For research purposes the computational intelligence available with the Web is still by and large an unexplored territory in social sciences (Stanton 1998; Birnbaum 2000; Englis and Solomon 2000; Reips 2000; Klassen and Jacobs 2001; Jespersen 2004).

Declining response rates – often termed ‘survey fatigues’ – is of major concern to research. The use of emails has made it easier to target respondents but also easier for respondents to opt-out of research studies (Stanton 1998; Reips 2000; Sheehan 2001). Collecting data with simulations by use of Web technology places the data collection methodology in a known medium for playing games and being active. Hence the appeal of behavioral simulations can be expected to be higher due to Web technology (Stanton 1998; Birnbaum 2000; Englis and Solomon 2000; Reips 2000; Klassen and Jacobs 2001; Jespersen 2004) visualizing the virtual decision-making process. Additionally, as games give players satisfaction from playing, so simulations will provide respondents with satisfaction of a problem-solving experience close to the real worlds of the respondents. Other possible reasons why the response rate may increase with Web technology are (i) the minimization of the response time for participants, (ii) the ease and flexibility of participation since the Internet is available twenty-four hours, seven days a week, and (iii) the lower cost for participants (Birnbaum 2000; Coderre and Mathieu 2004). Still, the experience from email surveys that response rates declines just as for postal surveys may indicate to researchers that they should be very careful about the design of the data collection on the Web. Hence, the design of a Web-based methodology has to be very user-oriented. If the design becomes too heavy or too boring, then participants will drop out as they do on regular web pages.

Another important data collection issue is item completion. Here the Web technology is powerful. Item completion can be simplified through the use of pull-down menus, check boxes, radio button scales, and drop-down scales. These options really make the web-based approach user-friendly. Furthermore, the pages can be programmed to check whether all items have been completed before giving access to the next page, thereby eliminating missing values. Additional four advantages of the web technology can be listed: (i) it is easy to modify the research instrument or to create multiple experimental versions, (ii) the automated data collection eliminates coding errors, (iii) reduction of experimenter effects, and (iv) the ability to reach a larger and more diverse subject pool. But also, the following disadvantages have appeared: (i) the potential for systematic bias and measurement error and (ii) the unwillingness of respondents to provide sensitive information over what they view as insecure lines (Birnbaum 2000), as well as iii) the issue of technical variance among respondents and iv) the issue of a self-selected sample (Reips 2000). Thus, as with any data collection methodology care and consideration should be demonstrated in order to maintain data validity.

4. Ensuring the validity of data

For Web technology supported simulations to become a new methodological trait, the data collected by means of this new research strategy are to be proved valid. Applying Web technology to the data collection vehicle raises the issue of convergent validity compared to a laboratory based study. Comparing laboratory and Web versions of surveys, scales and experimental variables, Krantz and Dalal (2000) find a remarkably consistent correspondence between the two, thereby proving the validity of Web technology as part of research strategies. A notion supported by Coderre and Mathieu (2004) finding that the predictive power of information gathered by email survey outperformed that observed with postal or telephone collected information.

Another issue of internal validity is construct validity by which the results of the data collected with simulations follow theoretically predicted trends. Paying attention to the details when creating the virtual world of the simulation will ensure construct validity. Research using agent-based computational modeling proves that this is not a critical issue for simulations though much lies with the
thoroughness of the researcher (Wahle, Bazzan et al. 2002; Boer, Ebben et al. 2003; Chaturvedi, Mehta et al. 2004).

The external validity of data collected with simulations is very critical for the generalization of the research results. Despite the many benefits of a simulation it is a constructed reality that the respondents enter. This constrains the analysis results and introduces the possibility of explaining nothing but the behavior within the simulation. To counter this criticism several validation questions should be designed into the simulation to observe whether the simulated managerial decision-making process resembles the actual decision-making processes in the participating companies. Still, it must be expected that some results can be explained by the behavior of the respondents being influenced by the simulation as are decision-makers in specific decision-making situations in companies. Furthermore, to validate data, knowledge of various decision-making characteristics should be obtained from the respondents using a questionnaire mapping relevant environmental and company moderators of the behavior of the respondents.

Combining simulations and Web technology introduces a new type of validity known in agent-based modeling as solution validity (Chaturvedi, Mehta et al. 2004; Jespersen 2004). Solution validity ensures the compatibility between the empirical data collection methods and the model solution in order to secure reliable and valid data. This means restraining the technological possibilities. Though a model of a real decision-process, some constraints have to be imposed on Web technology. A simple example is the habitually used back and forward buttons on the Internet. If these are not locked, the respondents would be able to go back and alter decisions without the researcher’s knowledge. Such a simple design consideration would jeopardize the data validity.

The data validity issue is always critical when collecting data but the dimensions to consider are increased when Web technology is applied and several methodologies are combined into one. Still, behavioral simulations will not jeopardize the validity of data collected, if designed properly.

5. Info@performance.NPD

The behavioral simulation as the data collection instrument example was developed for a study of new product development decision-making (Jespersen 2004). The aim was to gain greater insights concerning the value of information when new products are screened in the various phases of the new product development process. The analysis of the information value information two-dimensional including both the value-for-money and the decision-value of information. Hence, the behavioral simulation focused on the information processing cycle of decision-makers, i.e. whether information was acquired, and whether the acquired information was used for the screening decisions.

5.1 The model: A virtual NPD process

A virtual NPD process was created in which the participants were asked to acquire information for the evaluation of a new product idea as the product idea moved through the NPD process from idea to market planning. Hence, the simulation ended with the decision whether to launch or not. Because the game scenario used the Internet as medium, it was named Info@performance.NPD. Figure 1 depicts the structure of the simulation.

Combining simulations and Web technology introduces a new type of validity known in agent-based modeling as solution validity (Chaturvedi, Mehta et al. 2004; Jespersen 2004). Solution validity ensures the compatibility between the empirical data collection methods and the model solution in order to secure reliable and valid data. This means restraining the technological possibilities. Though a model of a real decision-process, some constraints have to be imposed on Web technology. A simple example is the habitually used back and forward buttons on the Internet. If these are not locked, the respondents would be able to go back and alter decisions without the researcher’s knowledge. Such a simple design consideration would jeopardize the data validity.

The data validity issue is always critical when collecting data but the dimensions to consider are increased when Web technology is applied and several methodologies are combined into one. Still, behavioral simulations will not jeopardize the validity of data collected, if designed properly.

Figure 1: The simulation structure

The starting questionnaire contained questions on new product development in the companies related to market orientation, strategy, budget, new product evaluation and information use. Then, the participants were asked to play the role of the new product development manager in the company MacVic\(^1\) and were given a job description. Here after followed a brief account of the new product development situation in MacVic followed together with an introduction of the virtual company MacVic. The description of MacVic focused on the external

---

\(^1\) MacVic is a fully imaginary company invented for the simulation.
information, secondary technical information, phase of the NPD process, but always (Booz, Allen et al. 1968; Cooper and Kleinsmidt 1987; Souder 1989; Montoya-Weiss and Calantone 1994; Brown and Kleinsmidt 1995; Crawford 1997; Benedetto 1999; Cooper 1999; Henard and Szymanski 2001). Hence, we aimed to ensure that the participants were given a monetary budget which could not be exceeded. The size and use of the budget ensured that the participants had to decide among the different information products, since it was only large enough to cover two-thirds of all the information products in the simulation. The size and use of the budget were validated with a cross analysis of questions from the starting questionnaire.

The simulated NPD process in the game scenario contained a thorough description of the status of the product idea in the five phases of the development process (idea, concept, prototype, product, and market planning) along with a total of thirty-six information products available at different points in the product development process. The available information products that the participants could acquire were designed in accordance with the 'best practices' guidelines developed by research on NPD success and failure (Booz, Allen et al. 1968; Cooper and Kleinsmidt 1987; Souder 1989; Montoya-Weiss and Calantone 1994; Brown and Eisenhardt 1995; Cooper and Kleinsmidt 1995; Crawford 1997; Benedetto 1999; Cooper 1999; Henard and Szymanski 2001). Hence the information products were different in each phase of the NPD process, but always represented the four information types - secondary market information, primary market information, secondary technical information, primary technical information. Furthermore, the information products were ensured a realistic touch through discussions of content, format and costs with professional market research companies. Additionally, the level of competitive stress facing MacVic and as such the participants were designed experimentally and regulated at three levels - low, medium, and high. The participants would encounter competitive stress in the concept and the prototype phases. To measure whether the acquired information was useful to the decision-making each phase ended with the participants evaluating the potential of the new product idea in the simulation on a scale from zero to one hundred.

To illustrate the game scenario - Info@performance.NPD, figure 2 shows the concept phase of the NPD process in the simulation. From these web-pages it should be possible to get an idea of how the simulation worked. The first Web-page is a description of the product development in the concept phase (picture 1) - what had happen to the product idea since the idea phase. Then the participant can choose from a list of information products (picture 2). The participant rates each information product on an importance scale. The participant then receives the ordered product (pictures 3-4), and rates the value of the individual information product (picture 5). Finally, the participant is asked to evaluate the product idea (picture 6).
5.2 Sampling

The sample constitutes larger International companies having their R&D unit placed in Denmark. The targeted companies develop either high or low technology products for the consumer market. Low technology products were represented by the food industry whereas high technology products were represented more diversely by industries such as telecommunication, personal computers, kitchen hardware, speakers, washing machines, tumblers, and headphones. The selection criterion was that the company did product development of consumer products in Denmark. This reduced the population of high technology companies tremendously, and introduced another constraint on the low technology companies: The food companies had to be of a reasonable size because primarily large high technology companies are performing new product development in Denmark. The participants were found either on basis of their job title as R&D manager/director or through organograms from which it was possible to determine whether it would be relevant to contact the marketing, engineering or project manager. A total of forty-two companies used the simulation.

5.3 Data validity

Two important aspects of the simulation as a data collection methodology are data validity and the judgment by the participating companies. Data validity was found through analyses of i) the budget, which showed that the participants used the same amount of money in the simulation as they would have in their companies, ii) the price of the different information products, where the relation between importance and the acquisition of an information product demonstrated a reasonable pricing in the simulation, and iii) the information utility of the individual information products. With high average ratings for all information products on the dimensions – relevance, quality and novelty – the information products did not obscure the decision-making process. The participants evaluated the simulation on four dimensions – coherent and easily comprehended new product development process, realistic content, exiting participation, and time consumption (see figure 3).

Figure 3: Appeal of the behavioral simulation

In general the simulation was found fun to do and very realistic. The only negative aspect was as expected the time used for participation. Though the measured time for participation was sixty minutes with the possibility to break this time down in minor intervals, the reaction to time from the participants was expected as time is always critical. Despite of this several participants expressed their positive experience of science going new ways. All of which seems promising for behavioral simulations as a new methodological trait.

6. Implications and future trends

A behavioral simulation is a suitable methodology for the collection of real-time data in all fields as it provides high interaction with respondents and immediate pay-back
through learning for the time invested in participation. Furthermore simulations appeal to respondents by applying Web technology. Also the combination of several methodological traditions is found beneficial for future methodological developments as this will also encourage new collaborations among research fields thereby enhancing the research contribution of future research.

Of course behavioral simulations are to be used with extreme care to both design and sampling in order to ensure validity as well as generalizability. Also the speedy technological development imposes a great challenge on researchers and as such the behavioral simulation can only be expected to be a continuous development to preserve its high appeal.

References


Crawford, C. M. (1997). New products management, IRWIN.


Hare, M. and P. Deadman (2004). "Further towards a taxonomy of agent-based simulation models in environmental