Editorial for the ECRM Special Issue of EJBRM 2012

December 2012
Editor: Ann Brown
Guest Reviewers: Martin Rich, Barbara Crump, Marie Ashwin and Angela Benson

Editorial to the conference issues

The subject of research methods in business is showing an extra-ordinary level of activity and innovation and this conference (the 11th European Conference on Research Methods in Business and Management) reflected this. These papers dealt with the problems facing management researchers in a variety of ways. Many Papers offer help in applying new methods such as Design Science. Perhaps the outstanding feature was the number of papers that concentrated on the assessment of research methods and the quality of results achieved by each type.

The final selection of papers was made by the editor of the Journal, who is grateful for the help provided by the guest reviewers. The papers selected were chosen for their quality of writing, their relevance to the Journal’s objective of publishing papers that offer new insights or practical help in the application of research methods in business research and the degree of innovation in the subject matter.

The chosen Papers

The first two papers critically assess the current position of research methods from separate view points. One (Stacy and Stacy) reviews the change in attitudes to ethical standards applied in research over the last century and suggests that current standards need reassessing. The second paper (Venables and Baskerville) seeks to use Design Science to generate rules for the assessment of competing research methodologies.

In the paper on multi-methodological methods (Marais) it is argued that the need for using a combination of research methods for individual research projects will increase as we face ever more complex socially relevant problems. Marais develops a 3 Dimensional framework (incorporating the key choices of; the key components of a project, the criteria required to produce high quality empirical work, the goal of the research project) to help researchers recognise when and how to adopt multi-methods.

A highly professional assessment is given by Cardoso and Ramos of how well the principles for conducting and evaluating interpretive IS research (proposed by Klein and Myers (1999) in their landmark paper in the MIS Quarterly), have been applied in subsequent papers in the two most reputed journals in IS research – MIS Quarterly and ISR. They find mixed results!

Design Science seems to be acquiring more supporters – particularly for research into Information Systems. The paper by Gacenga et al uses the technique to focus on the design process.

Two of the papers offer specific support for the application of empirical research methods – one on interview techniques for qualitative research (Hirose et al) and the second on web surveys (Bataard). Hirose et al introduce a highly original approach, taking the view that the emotional state of the interviewee is of great significance to the accuracy and quality of qualitative interviews. They propose that existing sensing devices can be used to assess the emotional responses of interviewees and hence determine true responses more effectively. Bataard offers us an excellent guide to developing effective web surveys drawing on the existing theory developed for surveys carried out by traditional methods.

The conference received a good number of papers on the teaching of research methods. This issue includes two of these papers. The thorny issue of evaluating the value of teaching courses in higher education is addressed in the first paper (Griggs et al). The paper argues that it is the learning outcomes that are of importance and that there are several stakeholders for which this is of interest. Their approach is to propose a balanced score card framework for evaluation. The final paper (Mkansi and Acheampong) deals with the complexity of the choice of research philosophy that all PhD students face at the beginning of their research. The paper calls for the development of a ‘planned, systematic framework and procedure’ to help students to make sense of the many options available to them.
Integrating Sustainable Development into Research Ethics Protocols

Anthony Stacey¹ and Julie Stacey²
¹Graduate School of Business Administration, University of the Witwatersrand, Johannesburg, South Africa
²Centre for Sustainability in Mining and Industry, University of the Witwatersrand, Johannesburg, South Africa

Abstract: The challenges and crises that face organisations are frequently the result of unintended, unanticipated or unforeseen consequences of well-intended decisions. In this paper the role of research ethics is analysed in as far as it contributes to or militates against future decision-outcomes that compromise the development of sustainable societies. The concept of sustainable development (Brundtland, 1987) came into being because of the enormous environmental, social and economic challenges manifesting in contemporary life. Many of these challenges are unintended consequences of historical decisions which were not necessarily considered unethical at the time. However, ethical standards are aligned with the norms and values upheld by societies and therefore change over time. There are various historical examples of research that would not be contemplated or permitted today due to contemporary ethical standards. This raises a question as to the frontier ethical standards against which contemporary research will be judged in the future. Research and innovation frequently have unintended or unanticipated consequences, which may be either favourable or detrimental, and outside the scope of the research. Various recent examples of negative unintended consequences of research are cited. It is shown that contemporary research ethics protocols do not typically draw attention to the possibilities of negative consequences due to their almost exclusive focus on the human participants and the research process itself. Using the five capitals model of sustainable development (natural, human, social, manufactured, and financial) it is demonstrated that typically the unanticipated consequences of research occur in one or more of the capitals other than the capital in which the research is being carried out. The central tenet of the paper is that researchers ought to consider and be accountable for favourable and detrimental consequences of their research. To this end, a taxonomy of negative consequences is developed with recommendations as to the researchers’ obligations in respect of the unanticipated consequences of their research, classified as unaddressed, unintended, unexpected, unforeseen and unforeseeable. A more comprehensive view needs to be taken of the ethics of business and management research to include procedural ethics, intrinsic ethics, and extrinsic ethics. Sustainable development is integrated into research ethics by considering the interdependencies of the five capitals of sustainable development and proposing a protocol based on the twelve features of sustainable societies. It is argued that evaluation of research using the proposed ethics protocol would militate against or mitigate the negative unanticipated consequences of research, thereby contributing to more sustainable societies. It may be argued that individuals cannot be held responsible or accountable for that which is unintended. However, by applying their minds using a more powerful ethical framework, researchers are more likely to anticipate and mitigate against these potentially latent and often long term impacts of their research. The approach in this paper is practical and applied rather than theoretical, philosophical or moral; business ethics and the ethics of medical research involving human or animals are excluded.

Keywords: extrinsic research ethics, five capitals; sustainable society; business decisions

1. Introduction

The concept of sustainable development (Brundtland, 1987) came into being because of the enormous environmental, social and economic challenges manifesting in contemporary life. Due to its very nature, sustainable development has pervaded all aspects of society – including business – in recent decades. Managers of organisations have been challenged to apply their minds, modify business practices and potentially amend their worldview based on this dynamic and maturing understanding of sustainable development. It therefore behoves all academics and professionals in business and management to consider their activities and the ethics of their research in the context of this paradigm.

There are many theoretical perspectives from which ethics can be viewed, including philosophical, legal, historical, moral, cultural, and human rights. However, research ethics will be addressed from a broad practical and applied perspective in this paper. It will be shown using historical examples that research ethics are dynamic and evolutionary. It can therefore be deduced that the ethical standards against which contemporary research is being judged are unlikely to be the same as those which will...
be applied by future researchers, who will apply profoundly different social, professional and academic values. Further, the insufficiency of contemporary research ethics will be demonstrated by showing that unanticipated negative consequences of particular historical research would not have been mitigated by the application of contemporary research ethics.

The paper concludes by suggesting that an ethical protocol based on the twelve features of a sustainable society will narrow the gap between current research ethics protocols and the standards by which contemporary research will likely be judged by future generations.

2. Progression of research ethics

A number of historical research projects have achieved notoriety on account of their methodologies or unanticipated consequences. Some of these have been well-publicised, and have been dubbed “Mad Science” by Schneider (2009).

The experiments involving “Little Albert” (Watson & Rayner, 1920) are controversial for a variety of reasons. In these experiments, the reactions of a nine month old boy to a variety of animate and inanimate objects were observed. Thereafter, the researchers conditioned the infant to fear a white rat by associating the rat with a loud clanging noise, resulting in the boy crying at the sight of the white rat (even when the noise was absent). The manner in which the experiments were carried out and the analysis reported have been the subject of criticism by subsequent authors (Harris, 1979; Paul & Blumenthal, 1989). Deliberately invoking fear in an infant in a laboratory setting, with or without parental consent, would not be contemplated by contemporary social scientists.

The Tuskegee Syphilis Study by the United States Public Health Service which began in the early 1930’s involved withholding of medical treatment from a racially distinct group of syphilis infected individuals (Thomas & Quinn, 1991). Such was the affront to the dignity of not only the participants, but also an entire demographic group, that a presidential apology was offered by Bill Clinton on behalf of the United States government in 1997 (Harter, Stephens, & Japp, 2000). The study involved taking advantage of vulnerable individuals and is highly unethical in a number of respects (Experiment-Resources.com, 2008).

The so-called “Milgram experiment” (Milgram, 1963) purported to be an experiment to determine whether the administering of increasingly painful electric shocks by a volunteer research assistant (the “teacher”) would improve the memory of the volunteer research subject (the “learner”). In fact, it was the “teacher” that was the actual subject of the research, and the “learner” was receiving no electric shock whatsoever and was merely an actor in collusion with the researchers. The purpose of the research was to evaluate the subjects’ compliance with commands given to them by an authority figure. The findings were that, in response to authority, about two thirds of the “teachers” administered the maximum possible electric shock to the “learners” despite the extreme pain that the “learner” appeared to be in. The problem was that the experiment was extremely stressful and traumatising for the “teachers” and it is generally agreed that such an experiment would never receive ethical approval today (McArthur, 2009).

Having reviewed some examples of historical research in the social sciences, it is evident that, while these experiments may have been regarded as ethical at the time of the research, their methodologies would have been rejected when judged by contemporary ethical standards. Research ethics clearly evolve in order to remain consistent with societal values. Similarly, businesses have “contracts” with society that evolve from frontier expectations through semi-formal contracts, to formalised contracts such as legislation (Bonini, Mendonca, & Oppenheim, 2006). Therefore, contemporary professional and academic researchers in business and management studies should ask the question: Against what frontier ethical expectations could current business and management research be judged by future generations?

3. Insufficiency of contemporary research ethics

The emergence of contemporary research ethics has not been a simple linear process (Isreal & Hay, 2006). Nevertheless, many of the generally accepted principles of research involving human participants can be traced back to The Nuremberg Code (1949). Although this code, together with the Helsinki Declaration (World Medical Association, 1964) are directed towards the ethics of medical research involving human participants and material, they are the seminal works of contemporary research ethics in the social sciences, including business and management.
The following principles are common to contemporary research ethics protocols and standards:

- Informed consent, which implies the avoidance of overt or secret participant observation (Bulmer, 1982).
- Privacy of participants (confidentiality and anonymity).
- Avoiding harm (including psychological effect) and doing good.
- Cognisance of vulnerable groups.
- Participants’ right to withdraw or terminate.
- Restricted use of data.
- Due care in the storage of data.
- Avoidance of conflicts of interest.

Reviewing some of social science’s most notorious research, it may be argued that the unanticipated consequences of the research could have been avoided if the research had met contemporary ethical standards. However, conforming to such a protocol would not necessarily have been a guarantee that unanticipated consequences would not manifest.

In various publications (P. Duesberg, 1992; P. Duesberg, Koehnlein, & Rasnick, 2003; P. Duesberg & Rasnick, 1998; P. H. Duesberg, 1991) Duesberg and his colleagues challenged the generally accepted understanding that HIV causes AIDS. Duesberg and Rasnick (1998) argued that AIDS is attributable to drug use, not HIV. They conclude that “the AIDS dilemma could be solved by banning anti-HIV drugs, and by pointing out that drugs cause AIDS.” Butler isolates the two competing paradigms that dominated the HIV/AIDS policy debate in South Africa, which he labelled the “mobilization / biomedical” and the “nationalist / ameliorative” paradigms (Butler, 2005, p. 592). He goes on to highlight that each of these paradigms was premised upon different assumptions about the challenges that HIV/AIDS posed. What has become known as the Duesberg hypothesis had the support of a community of so-called AIDS-denialists, who were influential in the formulation of the South African government’s post-apartheid HIV/AIDS policy. It is estimated that the loss of more than 300 000 lives could have been averted were it not for this policy (Chigwedere, Seage, Gruskin, Lee, & Essex, 2008; Nattrass, 2008). Despite the weight of research that has subsequently criticised and refuted their hypothesis, Duesberg and his colleagues did not appear to have breached contemporary ethical standards in the course of their research.

This narrow ethical focus is not limited to the social sciences. Research by economists has illustrated the benefits and incentives of introducing financial intermediaries between borrowers and lenders (Diamond, 1984; Leland & Pyle, 1977). In terms of this “originate-to-distribute” model of banking, banks (the originators) lend money to borrowers, and then aggregate the debt and sell it on to third party lenders (Purnanandam, 2010). Prior to the global financial crisis that occurred late in 2007, banks in the United States had adopted this originate-to-distribute model of granting credit, specifically mortgage loans. There are benefits of this model of banking to the borrowers, the banks and the lenders such as enhancing the liquidity of the secondary syndicated loan market (Berndt & Gupta, 2009). However, Brunnermeier (2009) shows that the financial crisis was partially attributable to the originate-to-distribute banking model. Although there was nothing unethical from a research perspective, the liquidity crunch and financial crisis were an unintended consequence of this financial innovation (Brunnermeier, 2009). As another commentator observed “… what [the financial industry] did was to innovate itself, and the rest of us, into a big, nasty mess” (Krugman, 2007).

It is clear from the summary by Aksin, Armony, and Mehrotra (2007) that there has been extensive research into optimisation of capacity planning, queuing, and personnel scheduling for call centres. Optimisation strategies have undoubtedly been adopted that had negative impacts on the customer service representatives. At face value, the Sacrificial Human Resources Strategy (Wallace, Eagleson, & Walderssee, 2000) could be considered callous in its disregard for the impact on individuals and society. A parallel field of research exists that looks into the human resource issues in call centres (Aksin et al., 2007) in response to the impact that has not been studied in the traditional call centre research.

Researchers may also face the possibility of their findings being used or applied in ways which they had not intended. Mastrandrea and Schneider (2004) identified this possibility in the highly
controversial and politicised field of climate change when they observed that “… users of scientific results (including policy-makers) will undoubtedly make their own assumptions about the probability of different outcomes, possibly in ways that the original authors did not intend” (Mastrandrea & Schneider, 2004, p. 571). Contemporary research ethics protocols are silent on any obligation on researchers to consider the possibility of (mis)application, (mis)interpretation or (ab)use of their research by others. The ethical principles of informed consent, privacy of participants, avoiding harm and doing good, cognisance of vulnerable groups, participants’ right to withdraw or terminate, restricted use and due care in the storage of data, and avoidance of conflicts of interest apply to researchers’ own objectives and interactions, but do not address the intentions and purposes of third parties who may not uphold similar ethical standards.

It is clear that, no matter how rigorously and meticulously researchers adhere to contemporary research ethics protocols, these protocols do not necessarily militate against negative consequences. Therefore, it is argued that current research protocols are insufficient to adequately protect society at large from the totality of consequences of research, or to contribute with some degree of certainty towards sustainable societies. It is neither feasible nor logical to expect that all potential risks can be eliminated (i.e. reduced to a probability or consequence of zero); there remains nonetheless a gap between contemporary research ethics and the protocols needed to promote and support sustainable societies.

4. Unanticipated consequences and unsustainability

The Five Capitals Framework of sustainable development (Porritt, 2007) has been developed as a basis of understanding the entirety of sustainability using the concept of wealth or capital as it is understood in economics, and is encapsulated in Table 1. This model is a constructive framework for analysing the diverse consequences of research activities that to date have not been explicitly addressed, mitigated or avoided by researchers.

Table 1: The five capitals framework (Porritt, 2007, p. 139)

<table>
<thead>
<tr>
<th>Capital</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>Any stock or flow of energy and matter that yields valuable goods and services. It falls into several categories: resources, some of which are renewable (timber, grain, fish and water), while others are not (fossil fuels); sinks which absorb, neutralize or recycle waste; and services, such as climate regulation. Natural capital is the basis not only of production but of life itself.</td>
</tr>
<tr>
<td>Human</td>
<td>Health, knowledge, skills and motivation (all of which are required for productive work), as well as individual’s emotional and spiritual capacities, Enhancing human capital (for instance through investment in education and training) is central to a flourishing economy.</td>
</tr>
<tr>
<td>Social</td>
<td>Structured, institutions, networks and relationships which enable individuals to maintain and develop their human capital in partnership with others, and to be more productive when working together than is isolation. It includes families, communities, businesses, trade unions, voluntary organisations, legal / political systems and educational and health bodies.</td>
</tr>
<tr>
<td>Manufactured</td>
<td>Material goods – tools, machines, buildings and other forms of infrastructure – which contribute to the production process but do not become embodied in output.</td>
</tr>
<tr>
<td>Financial</td>
<td>Plays an important role in our economy by reflecting the productive power of the other types of capital, and enabling them to be owned and traded. However, unlike the other types of capital, it has no intrinsic value; whether in shares, bonds or banknotes, its value is purely representative of natural, human, social or manufactured capital.</td>
</tr>
</tbody>
</table>

The unit of analysis in the AIDS research by Duesberg and Rasnick (1998) was the health and behaviour of the individual, which falls within the definition of human capital. However, a substantial consequence of the research was in the political and social spheres of social capital. The implementation of the originate-to-distribute banking model based on much earlier research (Diamond, 1984; Leland & Pyle, 1977) was an innovation within the domain of financial capital. The negative impact of the consequent financial crisis was a devaluation of not only financial capital but also human and social capital. The financial crisis potentially affected natural capital in as far as the resultant organisational budget cuts often affect the ‘non-core’ management issues first (as environmental issues are often perceived in business). Typically financial models have been developed for the management of personnel resourcing and scheduling at call centres, omitting the profoundly negative human and social consequences. If the most important research question to study is the longitudinal relationship between the aforementioned sacrificial call centre strategy and
market performance, as suggested by Wallace, et al. (2000), then human and social capital is discounted. It is well understood that technological developments, even as far back in history as the industrial revolution in the mid-1700's, resulted in extraordinary growth in manufactured and financial capital. Yet, the profound impact on the natural environment and the negative effects on society represent extensive destruction of natural, social and human capital at rates that exceed their ability to recover. The nuclear age ushered in similar developments in manufactured capital, but the military applications of the technology, the production of extremely high risk waste and worker exposure to radiation doses has come at a high price in natural, social and human capital – “nuclear power is a Faustian bargain” (Williams & Cantelon, 1984, p. xi). The social and human impact of contemporary manufacturing technology is clearly evident in the much publicised turmoil and tragedies experienced by individuals working in electronics manufacturing services companies (Branigan, 2010; Chan & Pun, 2010; Ngai & Chan, 2012). These examples demonstrate that the negative consequences of research activities can be the antithesis of sustainable development.

This analysis suggests that although typically the enhancement of capital as a result of research or innovation occurs as intended within one of the five capitals, this may sometimes be achieved at the cost of destruction of value in one or more of the remaining capitals. Schienke et al. (2009) proposed that responsible conduct of scientific research recognises three distinct components of research ethics: procedural ethics, intrinsic ethics and extrinsic ethics. As per their definition, extrinsic research ethics are:

“Ethical issues extrinsic to the production of scientific research – that is, ethical issues in how the outcomes of science research impact society, such as: policy-making; lawsuits; changes in social norms; and education and entertainment.” (Schienke et al., 2009, p. 322)

This triad of ethical domains has been designated the Ethical Dimensions of Scientific Research (Tuana, 2010). By introducing the concept of extrinsic ethics, Schienke et al. (2009) and Tuana (2010) are effectively proposing that researchers should integrate and apply their minds to the capitals other than that in which the research is being carried out, especially regarding the research outcomes. There are examples of contemporary research into supply chain network design, that explicitly address issues pertaining to environmental capital (Wang, Lai, & Shi, 2011) and social capital (Pishvaee, Razmi, & Torabi, 2012). The collaborative research involving Indigenous communities in Canada (Ball & Janyst, 2008) is an example of the incorporation of social capital into the ethics of research. The principles of collaboration and consultation are now embodied in the document which prescribes ethical conduct for research involving humans in Canada (Brant Castellano & Reading, 2010; Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada, 2010; Flicker & Worthington, 2012). Similarly, the necessity to engage with affected racial and ethnic groups in specific research contexts is acknowledged in the United States (Eckstein, 2012) and engagement with Indigenous Australians is embodied in the principles of carrying out health research among these populations (Commonwealth of Australia, 2003; Jamieson et al., 2012). In the information and communication technology field, Bilandzic and Venable (2011) propose a methodology of “participatory action design research” for urban informatics and suggest that this methodology will “…enable closer collaboration between academic researchers and the communities that they serve and benefit.”

5. A taxonomy and hierarchy of negative consequences of research

The central tenet of this paper is that researchers are responsible for considering all the outcomes and consequences of their research, whether or not those consequences are favourable and whether or not they are within the immediate objectives of the research. It can be argued that there are certain unforeseeable consequences of research for which researchers cannot reasonably be expected to accept responsibility. Conversely it is argued that researchers cannot merely abdicate responsibility by expedient definitions of scope. Therefore a taxonomy of negative consequences of research is helpful in articulating these consequences, allied to the ethical obligations of researchers.

Table 2 gives a proposed taxonomy of negative consequences of research which are considered to be out of the research scope; the categories of consequences are ordered in descending order of probability. The nondisclosure or omission from publication of negative, non-target effects or side effects that are within the research scope (Holdrege, 2008) would be regarded as unethical in terms of current standards; this situation is therefore not addressed in Table 2. It is proposed that
Researchers’ ethical obligations vary according to the risk associated with the outcome, which is a function of the consequence and probability of the event. In terms of this taxonomy, researchers have some obligations for even the most problematic negative consequences – those that cannot be foreseen.

**Table 2: A proposed taxonomy and hierarchy of negative consequences of research**

<table>
<thead>
<tr>
<th>Categorisation</th>
<th>Definition</th>
<th>Probability</th>
<th>Researchers’ obligations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unaddressed</td>
<td>Consequences are known to the researchers but are not examined because of the predefined scope.</td>
<td>High</td>
<td>Public disclosure and engagement with stakeholders to achieve consensus on research scope and objectives.</td>
</tr>
<tr>
<td>Unintended</td>
<td>Consequences are known to the researchers, but with low enough probabilities that associated risks can be ignored.</td>
<td>Low</td>
<td>Include in the scope of the research measures to avoid negative consequences, despite the very low probability.</td>
</tr>
<tr>
<td>Unexpected</td>
<td>Consequences are known to the researchers, but with no anticipated risk because of zero probability of occurrence.</td>
<td>Zero</td>
<td>Disclosure and consultation within the professional and academic community to corroborate zero probability of occurrence.</td>
</tr>
<tr>
<td>Unforeseen</td>
<td>Consequences are unknown to the researchers, but could reasonably have been foreseen if researchers had applied their minds.</td>
<td>Unknown</td>
<td>Apply their minds to possible negative consequences, and their associated probabilities and impacts.</td>
</tr>
<tr>
<td>Unforeseeable</td>
<td>Consequences could not have been reasonably identified by researchers.</td>
<td>Undefined</td>
<td>Remain open-minded to possible negative consequences, and seek input on these, no matter which stakeholder(s) recognise(s) the risks.</td>
</tr>
</tbody>
</table>

Research is often mired in secrecy, and protected by copyright or patent laws, for reasons of economic competitiveness and gain. The systemic nature and scale of sustainable development challenges requires a worldview shift from protectionist to collaborative: hence the requirement for broader consultation. That this might require new research methodologies is acknowledged. Simultaneously, that “All people have the right to participate in and access information relating to the decision-making processes that affect their lives and well-being” is equally well entrenched in social and political doctrine through international law (United Nations Population Fund, 2012). This broader consultation could even result in unforeseeable consequences being serendipitously identified.

A similar taxonomy can be created for positive outcomes; ideally researchers would apply their minds to maximising positive outcomes towards sustainable futures. The obligations proposed in Table 2 are aligned with the typical hierarchy of risk treatment options as defined in ISO 31000:2009 (Purdy, 2010) which also apply equally to positive and negative outcomes. The preferential option is always to eliminate or avoid negative outcomes. By virtue of their activities, researchers are ideally positioned to identify and remove potential threats, and this should be their priority. It is worth noting here that the rigour of the research process necessitates proper observation and evaluation of activities. It has been argued (Hansson, 2011) that ethically problematic or risky activities should therefore only be carried out within a well-defined research context. However, a statement of responsibilities of researchers is worthless without a corresponding mechanism of accountability. The research ethics protocol is the appropriate means of holding researchers to account for applying their minds to all outcomes of their research.

### 6. A paradigm and protocol of research ethics for sustainable development

Tuana noted that “My point here is that this domain of Ethical Dimensions of Scientific Research is ripe for research informed by philosophy of science perspectives” (2010, p. 481). It is being advocated in this paper that the ethics of business and management research is ripe for transformation into the context of sustainable development, due to the interdependent and constantly interacting nature of the capitals on which organisational productivity depend. Organisations are after all the mechanism through which people transform natural and human capital into social, manufactured and economic wellbeing. The twelve features of a sustainable society (Forum for the Future, 2011) is a useful framework within which consideration of potentially negative research outcomes can be applied by researchers. It is noted that these statements are comprehensive, internally consistent and culturally neutral. A protocol corresponding to this framework should govern and guide research and
professional innovation activities. Table 3 gives a proposed research ethics protocol using the five capital framework based on the twelve features of sustainable societies. The proposed protocol is consistent with sustainable development principles of fair access to resources across generations (intergenerational equity), acknowledgement of the finite limits of the capitals, and the tension between needs and wants.

Table 3: Proposed research ethics protocol based on the twelve features of sustainable societies

<table>
<thead>
<tr>
<th>Construct</th>
<th>Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural capital</td>
<td><strong>Natural capital must be shared fairly between current and future generations – intergenerational equity. Natural capital processes and resources have finite capacities, and protracted timeframes may be required to realise these capacities.</strong></td>
</tr>
<tr>
<td>Harmful Effects of Extraction of Substances from the Earth.</td>
<td>Consider and describe how the research and its outcomes may influence the environment's capacity to disperse, absorb, recycle or otherwise neutralise the harmful effects (to humans and / or the environment) of the extraction and use of substances taken from the earth.</td>
</tr>
<tr>
<td>Harmful Effects of Artificial Substances.</td>
<td>Consider and describe how the research and its outcomes may influence the environment's capacity to disperse, absorb, recycle or otherwise neutralise the harmful effects (to humans and / or the environment) of the manufacture and use of artificial substances.</td>
</tr>
<tr>
<td>Ecological System Integrity and Biological Diversity</td>
<td>Consider and describe how the research and its outcomes may influence the capacity of the environment to provide ecological system integrity, biological diversity and productivity.</td>
</tr>
<tr>
<td>Human capital</td>
<td><strong>The needs of the many must enjoy priority over the wants of the few, for both current and future generations. The limited capacity of specific disadvantaged groups to engage must be addressed.</strong></td>
</tr>
<tr>
<td>Standard of Health.</td>
<td>Consider and describe how the research and its outcomes may influence the standard of health among individuals of all ages.</td>
</tr>
<tr>
<td>Relationships, Social Participation, Development and Learning.</td>
<td>Consider and describe how the research and its outcomes may influence individuals’ relationships and social participation, setting and achievement of personal standards of their development and learning throughout their life.</td>
</tr>
<tr>
<td>Work, Personal Creativity, and Recreation.</td>
<td>Consider and describe how the research and its outcomes may influence individuals’ access to varied and satisfying opportunities for work, personal creativity, and recreation.</td>
</tr>
<tr>
<td>Social capital</td>
<td><strong>Sustainable social capital requires at a minimum, the same preconditions as sustainable natural and human capital.</strong></td>
</tr>
<tr>
<td>Governance and Justice.</td>
<td>Consider and describe how the research and its outcomes may influence trust in and access to systems of governance and justice.</td>
</tr>
<tr>
<td>Positive Values and Sense of Purpose.</td>
<td>Consider and describe how the research and its outcomes may influence shared positive values and sense of purpose among communities and society at large.</td>
</tr>
<tr>
<td>Stewardship of Natural Resources and Development of People.</td>
<td>Consider and describe how the research and its outcomes may influence the structures and institutions of society to promote stewardship of natural resources and development of people.</td>
</tr>
<tr>
<td>Safe, Supportive Living and Working Environments.</td>
<td>Consider and describe how the research and its outcomes may influence the capacity for homes, communities or society at large to provide safe, supportive living and working environments.</td>
</tr>
<tr>
<td>Manufactured capital</td>
<td><strong>Limits of current technology and human knowledge must be acknowledged. Consideration must be given to needs versus wants, and the natural, human and social capital expense of creating new technologies, products and services.</strong></td>
</tr>
<tr>
<td>Use of Natural Resources, and Human Innovation and Skills.</td>
<td>Consider and describe how the research and its outcomes may influence infrastructure, technologies and processes to minimise the use of natural resources and maximise the use of human innovation and skills.</td>
</tr>
<tr>
<td>Financial capital</td>
<td><strong>The financial cost of human, social, natural and manufactured impacts for current and future generations that are external to the organisation must be acknowledged and disclosed.</strong></td>
</tr>
<tr>
<td>Accurate Representation of the Value of Capital.</td>
<td>Consider and describe how the research and its outcomes may influence the accuracy of the valuation of natural, human, social and manufactured capital in terms of financial capital.</td>
</tr>
</tbody>
</table>
One can speculate as to how the outcomes of the previously cited research activities may have differed with the application of the proposed research ethics protocol based on the twelve features of sustainable societies. The impact of HIV/AIDS in South Africa could have been lessened if there had been greater consideration and engagement with civil society by Duesberg and his colleagues regarding how the publication of their research might ultimately influence the standard of health among individuals, the development of people, and the capacity for communities and society at large to provide safe, supportive living and working environments, not least through influencing government policy and approach. The financial crisis might have been mitigated or militated altogether if research into financial intermediaries had been premised on the accurate valuation of natural, human, social and manufactured capital. Operations researchers may have applied different algorithms and objective criteria if they had explicitly considered how call centre optimisation could influence the health of customer service representatives, their relationships and their social participation. The frontiers of technological innovation would surely have been redirected had there been greater consideration for:

- the harmful effects of the extractive industries, and of manufacturing artificial substances,
- for ecological system integrity and biological diversity,
- for individuals’ relationships and social participation, and
- For shared positive values and sense of purpose among communities and society at large.

Mindful that the imposition of a bureaucracy of ethical procedures may be perceived to hamper academic research activity (Haggerty, 2004) the effect of the proposed protocol may rather be to increase the scope of such activities. Without the broader considerations embodied in the proposed protocol, the contribution of research to creating sustainable societies will be coincidental rather than by design.

7. Comments and conclusion

As part of a postgraduate personal development programme, students were required to assess their own research against the protocol proposed in this paper. Students were prompted to think systemically about the inputs, throughputs and outputs of their research, including to what use the results of their work could be put by ethical and unethical individuals. Although anecdotal, the results are nonetheless pertinent. The most common impact reported by students was the realisation that in conceptualising their research problem, their focus had been on the positive contributions to knowledge their work might achieve; they had not considered the potential negative consequences. When considering individual components of their research problem, some students discovered that certain aspects of their research could have unintended negative consequences, even when the overall outcome was likely to be positive. The feedback generally indicated that students would amend their final reports, the most extreme case being a student who decided to dedicate an entirely new chapter to discussion of the results of applying this framework.

Sustainable development as a concept provides a lens for both upside opportunity and downside risk research: the focus of this paper is the downside, or negative risk, given the scale of global challenges in each of the capitals. It is not possible to foresee every possible outcome of research. However, the current research ethic paradigm and associated protocols tend to encourage a narrow focus based on the rights of human participants, bounded exclusively by the scope of the research problem and methodology. The interconnected nature of the five capitals that support life and the systems conditions established through these relationships requires that researchers apply their minds to the broader features of sustainable societies when considering, at an ethical level, the possible outcomes of research. At the same time, the boundary conditions of limits, needs versus wants, and intergenerational equity must inform the ethical considerations of the research. The proposed protocol is the step change that is required to shift the current narrow research ethic paradigm to one that addresses the conditions for sustainable development. In order to move towards research that contributes to sustainable development, business and management researchers and professionals must apply a thorough understanding of systems thinking to their research in order to anticipate the full suite of its possible outcomes. No single researcher can be expected to have a complete understanding of all aspects of sustainable development: the proposed consultation, dependent on the significance of the possible outcomes, addresses this limitation. Societal heterogeneity (in many forms) further dictates that researchers engage publically on potential outcomes to better gauge the current status of social expectations. The use of the proposed protocol
will not necessarily change the outcomes of the research; the broader considerations advocated in this paper should encourage decision-making that is more likely to contribute to sustainable societies.

References


Hendrik Marais
Graduate School of Technology Management, University of Pretoria, Pretoria, South Africa
hc.marais@up.ac.za
marais.stnetwork@gmail.com

Abstract: The paper addresses the methodological commonalities linking quantitative and qualitative methodologies. It offers a three dimensional framework of research methodology that spans the assumed divide and shows that quantitative and qualitative research approaches can be mutually complementary and offer a more nuanced approximation of the truth. The framework, Research methodology in 3D (3D Framework), is based on a meta-methodological analysis and consists of three root dimensions. The first root dimension comprises the constituent components presented in all forms of research (concepts, questions/hypotheses, observation/measurement and communication). The second dimension of the framework comprises the core set of criteria that all scientific research must comply with (reliability, validity, ‘objectivity’ and replicability), each may be expressed under different labels and in different forms depending on the specific research tradition. The third dimension represents the basic goals of research, namely to describe, to explain or to understand. A three-dimensional framework emerges when the above three dimensions are combined. The framework shows some of the tangent planes between quantitative and qualitative scientific research and suggests possible complementarity between research approaches. The framework furthermore serves at least the following functions: firstly, as map to assist students in charting their way through a maze of new concepts. Secondly, the 3D Framework can serve as a useful planning instrument for comprehensive and complex research programmes. Thirdly, the framework can be utilized as a useful guide in evaluating research proposals on the one hand and research reports, on the other (an example of the latter is offered)

Keywords: research methodology, quantitative research, qualitative research, mixed methods

1. Introduction

According to Bryman, “combining quantitative and qualitative research has become unexceptional and unremarkable in recent years” (2006: 97). His content analysis of over 200 articles showed that the main reasons offered by authors for combining quantitative and qualitative research methodological approaches were ‘enhancement’ (to augment either quantitative or qualitative findings), sampling (to facilitate the sampling of respondents or cases) and triangulation (to cross-validate).

Against the background of Bryman’s observation, it would be interesting to identify research methodological preferences, if any, in Business and Management Studies. Since the European Conference on Research Methodology (ECRM) represented one the premier annual scholarly gatherings in Europe in these fields, it could be expected that these conferences should reflect methodological preferences. A relatively simple content analysis was consequently done of the papers presented at the most recent two conferences in 2011 and 2012 (see Ashwin, 2011 and McClean, 2012 for the proceedings). More specifically, the goal of the content analysis was to identify the relative popularity of qualitative, quantitative and multi-methodological approaches on the one hand and empirical and conceptual ones on the other (see the table note for the definitions). An associate of the author also coded a random sample (11) of the ECRM 11 papers to establish the reliability of the content analysis; she used the definitions of the different categories reflected in the note attached to Table 1. There was 91 percent agreement between the two sets of codings which was judged satisfactory given the relatively abstract nature of the categories. The results of the content analysis are summarised in Table 1.

An inspection of Table 1 shows several interesting aspects of research in business and management studies of which the following four are perhaps the most relevant from the perspective of the present paper. Firstly, over the two years of these two conferences qualitative empirical papers predominated, followed by quantitative empirical and qualitative conceptual papers. Secondly, conceptual papers represented 50 percent of all the papers presented at ECRM in Caen and Bolton. Thirdly, only two papers were presented that dealt with quantitative conceptual issues, i.e. theoretical and...
methodological underpinnings of research – as opposed to 25 or 17 percent that addressed qualitative conceptual issues. (Perhaps, one wonders, whether this can be attributed to the relatively long history of conceptual questions of a quantitative nature and what seems to be a more recent item on the qualitative agenda of methodologists.) Fourthly,

Table 1: Distribution of papers presented at ECRM 10 (2011) and 11 (2012)

<table>
<thead>
<tr>
<th>Category</th>
<th>Empirical papers</th>
<th>Conceptual papers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Qual (%)</td>
<td>Quan (%)</td>
<td>Multi (%)</td>
</tr>
<tr>
<td>Paper</td>
<td>19 (29)</td>
<td>19 (29)</td>
<td>4 (6)</td>
</tr>
<tr>
<td>PhD</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>Paper</td>
<td>17 (31)</td>
<td>13 (24)</td>
<td>4 (7)</td>
</tr>
<tr>
<td>PhD</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>15</td>
<td>7</td>
</tr>
</tbody>
</table>

Totals and Row Percentages of ECRM 10 + 11

| Total (%) | 48 (33) | 35 (24) | 11 (7) | 25 (17) | 2 (1) | 14 (10) | 9 (6) | 146 |

Notes:

1. **Empirical papers**: Reports on completed research projects involving some quantitative or qualitative form of observation/measurement was used for information gathering. **Conceptual papers**: Those that considered theoretical and methodological underpinnings of research or set out project plans (Plan) and did not report an empirical research project. **Qual**: Papers that dealt only with qualitative research; **Quan**: Papers that dealt only with quantitative research; **Multi**: Papers that covered both qualitative and quantitative empirical or conceptual research

2. Percentages have been rounded-off and up and do not necessarily add up to 100 percent.

Less than one fifth of the papers could be classified as being of a multi-methodological type (25 or 17 percent). Clearly, while multi-methodological research did receive a fair amount of attention, qualitative studies still predominated. In this regard, one is reminded of Blumberg, Cooper and Schindler’s comment, “Many scholars show a strong preference for either type of study. However, these preferences more likely reflect their own capabilities and experiences than a general idea about which type of research is more useful” (2008: 192). The present paper offers a three-dimensional framework of research methodology based on the assumption of complementarity rather than irreconcilability between different research approaches. This framework could further facilitate the incorporation of combinations of qualitative and quantitative research approaches to research and thus offer a more nuanced approximation of the truth. The paper, however, does not deny the value of good mono-methodological research irrespective of the methodological orientation or the important differences between the two approaches in, for example, the nature of the problem to be investigated, the researcher’s position in the research setting, theoretical orientation, research design, scope, information-gathering techniques and others (see, e.g., Denzin and Lincoln, 2005: 11–12).

There are many reasons why attention should be focused on complementarity between research approaches. Five methodological reasons include, firstly, the rationale underpinning the convergent and discriminant validation approach (e.g. Campbell and Fiske, 1958) and the parallel qualitative requirement of triangulation (e.g. Denzin, n.d.; Denzin and Lincoln, 2005) which highlights the benefits to validation of incorporating different methodological approaches into a study. A second justification for multiple methodologies is the growing importance of mixed-method designs as ‘a third methodological movement...as a pragmatic way of using the strengths of both movements’ [i.e. quantitative and qualitative], as Tashakorri and Teddlie described mixed methods (2003: ix). A third challenge to research methodology – and project management – emerges from the growing importance of multi-, inter- and especially transdisciplinary research (cf. Hadorn et al., 2008), as researchers address increasingly complex scientific and socially relevant problems, partly in response to the dynamics of science, technology and innovation policy, including international scientific collaboration programmes, and governments’ expectations that publicly funded research should show a positive return on investment by addressing national and even international problems (cf. OECD, 2007; Rosenfield, 2009). Fourthly, the possibility of constraints on public funding of universities in many countries, could stimulate multi- and interdisciplinary research as a strategy for ameliorating the
effects of cuts in research funds particularly in countries, such as South Africa, where some academic departments and disciplines have had to amalgamate. Finally, ‘MMRD [mixed methods] may dislodge ossified positions primarily maintained by identity politics associated with QL [qualitative] and QN [quantitative] methods’ to quote Bergman (2011: 275).

2. Main argument

The paper argues that national and institutional science policies increasingly challenge researchers to address complex socially relevant problems, and that such problems often require at least a multidisciplinary research design. The paper further argues that although caution is generally required in utilising different methodological approaches in the same study (see, e.g., Cameron, 2011, for some of the challenges), this should not necessarily inhibit researchers from venturing into the territory of multiple methodologies to find more comprehensive answers and solutions. It is further suggested that representations of quantitative and qualitative methodological approaches are often ideal types developed by philosophers of science and strong proponents of either side, but that they seldom occur in pure form in everyday research. Finally, the paper argues that there are indeed interfaces and sufficient commonalities between accountable methodological approaches to justify multiple methodological approaches, provided that this is done within an explicit methodological framework such as the 3D Framework proposed in this paper.

3. Objectives of the study

The main objective of the study was to develop a framework that could be used as an instrument to identify the interfaces between different methodological approaches and thereby utilise a broader spectrum of approaches than would otherwise be the case.

4. Information base of this paper

A brief note on the process that was followed in drafting this paper may be appropriate. The material, analyses, interpretations, conceptions and results reported in this paper are the result of critical reading of publications on research methodology, interactions with colleagues, active self-initiated and commissioned research, teaching of research methodology courses, research supervision, and research evaluation ranging from research proposals submitted for funding by a public research funding agency to completed research reports. In one sense, this paper represents a qualitative insider perspective on the subject, but in another sense it reflects a modest attempt at integrating in a fair way a wide spectrum of published knowledge on the process of research.

5. Conceptual framework

At the risk of offering redundant information, definitions of key concepts as used in this paper are provided:

- **Research**: The uncovering of the truth about a phenomenon (and/or its relationship with other phenomena) by means of scientific methods. The term ‘research’ as used in this paper equates to scientific research and includes qualitative research; it includes both theory building and testing.
- **Empirical research**: Research that relies on one or more forms of observation/measurement for gathering information, ranging from qualitative research (e.g. participant observation) to quantitative research (e.g. psycho-physiological indicators).
- **Science**: The accumulated published information generated by means of scientific methods, on the validity of which the peers in the particular field have reached consensus; the concepts ‘science’ and ‘scientific research’ cover all science cultures.
- **Social science**: The subset of science that focuses on human, social and cultural behaviour, including business and management studies.
- **Qualitative research**: Research that approaches phenomena from the perspective of the insider or subject in order to understanding the phenomena in their natural context. This approach uses qualitative ‘indicators’ such as words, stories, pictures and other communicative representations as non-numerical symbolic information on phenomena; its methodologies are normally less formalised, rigid, specific and explicated, but more comprehensively recorded.
- **Quantitative research**: Research that approaches phenomena from the perspective of the outsider in order to explain and predict the phenomenon under study in isolation. This approach uses numerical indicators of abstract concepts; its methodology is normally relatively formalised, rigid, cross-referenced and explicated, but more parsimoniously recorded by means of statistics.
6. Root dimensions of research methodology

To make comparisons between different methodological approaches requires the identification of a limited set of necessary and sufficient properties, attributes or characteristics of the scientific research process that could serve as the basis for comparison. From phenomenological and systems perspectives, and based on analyses of textbooks on research methodology, research publications and relatively wide research experience, a limited number of root dimensions that qualify research as ‘scientific’ were identified. These were further developed into a framework that would not define scientific research as a discreet point in the knowledge space – implying an all-or-nothing character – but would acknowledge that research can take a range of legitimate forms, in other words suggesting the existence and promotion of the utilisation of multiple approaches to generating scientific knowledge. A careful reading of methodological publications yields possible overlaps (e.g. Cresswell’s, 2003, three dimensions of knowledge claims, strategies of enquiry and methods), but also further conditions, although these can mostly be subsumed under one of the three root dimensions or are, upon closer inspection, particular to a very special research design. The root dimensions are summarised below, and a conceptual framework, Research Methodology in 3D (3D Framework), is offered in a later section.

Three root dimensions emerged from the conceptual analyses:

- **Constituent components or basic building blocks of the scientific research process:** concepts/constructs, hypotheses/theses, measurement/observation and communication of the research
- **Epistemic criteria that scientific research should comply with:** reliability/dependability, validity/credibility, ‘objectivity’/confirmability and replicability
- **The primary purposes for which research is normally undertaken:** description of a phenomenon and its relationship with other phenomena, explanation of causal relationships between phenomena, and in-depth understanding of phenomena.

6.1 Root dimension 1: Constituent components of the scientific research process

Most, if not all, human behaviour, ranging from everyday behaviour, to the arts and to scientific research, consists of at least the following four generic components (cf. e.g. Marx, 1963; Mouton, 1996):

- **Linguistic and para-linguistic representations of phenomena (words, concepts, constructs, symbols of all types):** Differences in the nature of concepts as used by qualitative and quantitative researchers have been debated by prominent scholars (see Mouton, 1996, for a useful summary), but it is important to note the obvious: concepts represent a necessary component in both qualitative and quantitative research traditions, although the labels and exact criteria may differ. In scientific research, concepts should meet the criterion of explication and operational specificity as far as possible.

- **Questions about phenomena (conjectures, theses, hypotheses):** Hypotheses (and their variants) serve to direct research and are key components of both qualitative and quantitative research. A hypothesis or thesis should be testable, and more specifically ‘rejectable’, to be useful in scientific research.

- **Observation of phenomena (sensory perception, experience, measurement):** Observation can range from participant observation to relatively sophisticated measurement of behavioural or even psycho-physiological responses. The criterion of control equates to ‘objectivity’, meaning that the researcher shows her/his autonomy in the observation process and is not merely ventriloquising.

- **Communication of what has been observed (general discussions, news reports, scholarly publications):** This component is seldom included as a basic component of research, but since the earlier definition of science refers to ‘published’ information on which ‘peers have reached consensus’, communication is elevated here as a key component of the scientific research process. This position is strengthened by the common observation that whether a particular research project qualifies as ‘scientific’, as well as its quality, are taken to be reflected by the journal – more specifically, the impact factor of the journal – in which the work was published or the conference at which it was presented!
The first three constituent components were originally summarised in this form by Marx (1963: 10) as being ‘basic elements of theory construction’. Specific criteria applicable to each of the components differentiate between forms of behaviour (e.g. distinguishing between a short story, an investigative journalism article and a research report). Each of the four components represents a continuum, and different research methodological approaches need not be, and usually are not, located at the same point on each of the continua (see the definitions in the previous section). The four components are shown in Figure 1 as continua, with the process criteria on the far right of each continuum and the arts on the extreme left of each continuum (cf. Marx, 1963: 11; Mouton and Marais, 1988: 157).

**Figure 1:** Constituent components and criteria of scientific research (adapted from Marx, 1963: 11)

The following brief notes on aspects of Figure 1 may be important in the context of this paper:

- The four constituent components are, of course, not independent of one another, and can perhaps be described as the results of a conceptual unfolding analysis (to use a concept originally coined by Coombs, 1964) of the core components of human behaviour in general and scientific research in particular.

- Furthermore, the representation offered in Figure 1 assumes that scientific research is a special form of human behaviour, the main difference between it and other forms of human behaviour (e.g. everyday conversations, different forms of art, and investigative journalism) being the specific content given to the italicised criteria in that figure. Also, the representation offered in Figure 1 allows for a range of social research methodological approaches – in general, qualitative research such as phenomenological studies would tend to be located to the left of experimental psychological studies, which would normally be closer to the extreme right of the continuum.

- Each of the components is expressed as a continuum – rather than an absolute point in some idealistic semantic space – with the obvious implication that different research methodological approaches could be located along the various continua, reminding one of Cresswell’s (2003) conclusion that qualitative and quantitative research are located on a continuum. In Figure 1, physics is theoretically positioned at the far right as the ‘ideal’, but this does not preclude, for example, history somewhere to the left and closer to the middle of the continuum.

- The key criterion for scientific research is italicised. Implicitly or explicitly, scientific research strives to be as specific as possible with regard to its concepts, to deal with testable research questions/hypotheses, not to engage in random observation or measurement, and to accept the validity of that on which peers in a particular research field have reached consensus.

- Generally speaking, and with reference to good research, experimental research (especially laboratory research) would be closer to the right extreme of the four components, while qualitative research would tend to be located somewhere to the right of the midpoint of each of the continua.
Scientific research takes place when these four components, each complying with the scientific criterion, interact in a systematic way.

6.2 Root dimension 2: Epistemic criteria that scientific research should comply with

The second root dimension of the framework is represented by the core set of criteria that all research, whether quantitative or qualitative, must comply with in order to qualify as scientific, as defined in section 5. Figure 1 reflects a process criterion for each of the four components of scientific research, namely: concepts should be operationalised as specifically as possible, preferably in terms of empirical referents; hypotheses should be testable; observation should be as controlled as possible; and, finally the communication of the findings should be subjected to peer review. These process criteria, however, do not specify the actual quality required of scientific research as the search for truth. For this, one should drill deeper into the nature of the scientific research process, as reflected in methodology textbooks and courses, as well as criteria applied to the assessment of dissertations/theses, journal articles, and conference papers. To supplement the author’s own conceptions and research practices, selected textbooks on research methodology for both quantitative and qualitative research were scanned to identify the primary criteria for scientific research (e.g. Alasuutari, Bickman, and Branner, 2008; Babbie and Mouton, 2001; Blumberg, Cooper and Schindler, 2008; Mouton, 1996; Mouton and Marais, 1988; Remenyi, Williams, Money and Swartz, 1998; see Marais, Pienaar-Marais and Gathua, 2011, for further sources consulted). These and other textbooks generally agree on two important issues. Firstly, both quantitative and qualitative methodologies subscribe to the same core set of epistemic criteria, secondly, different terms are used by proponents of those methodologies to label the criteria and, thirdly, the operationalisation of those criteria might differ (see Table 1). The criteria are, however, in the words of Babbie and Mouton (2001: 276), ‘all bootstrap conceptions of a sort. Although we should strive with everything in our power to do truly valid, reliable, and objective studies, the reality is that we are never able to attain this completely’, but, one might add, should pursue this in all research undertaken and explicitly report on it in all presentations. Table 2 offers an overview of the key criteria, using ‘bilingual’ nomenclature for the criteria, definitions and indicators, as normally used by quantitative and qualitative researchers respectively.

Table 2: Key epistemic criteria

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Quantitative research</th>
<th>Qualitative research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>Same findings upon replication?</td>
<td>Similat context yields similar findings?</td>
</tr>
<tr>
<td>Test-retest &amp; interrater reliability</td>
<td></td>
<td>Inquiry audit</td>
</tr>
<tr>
<td>Internal validity</td>
<td>Measured what intention was?</td>
<td>Compatibility between respondents’ and reported perceptions?</td>
</tr>
<tr>
<td>Experimental control; statistical triangulation</td>
<td></td>
<td>Prolonged engagement; member checks; quality record; narrative triangulation</td>
</tr>
<tr>
<td>External validity</td>
<td>Generalisability to population?</td>
<td>Applicable to other cases and contexts?</td>
</tr>
<tr>
<td>Random sampling</td>
<td></td>
<td>Purposive sampling; detailed descriptions of process</td>
</tr>
<tr>
<td>‘Objectivity’</td>
<td>Reflecting own views?</td>
<td>Findings not function of biases of researcher?</td>
</tr>
<tr>
<td>Control over subjective factors</td>
<td></td>
<td>Audit trail; trust &amp; rapport with subject; intersubjectivity</td>
</tr>
<tr>
<td>Replicability</td>
<td>Can next researcher replicate the study? Peer reviewed publication</td>
<td>Replicability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clear description of procedures?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appropriate peer-reviewed publication</td>
</tr>
</tbody>
</table>

Table 2 highlights, four main points. In the first place, the epistemic criteria, normally associated with quantitative research, have their equivalents in qualitative research (also see Golafshani’s, 2003, overview). It would be reasonable to define each of the criterion continua as unidimensional. In other words, each of the four epistemic criteria can be used to assess the quality of research irrespective of whether it is quantitative or qualitative. In the second place, notwithstanding the ‘universality’ of the epistemic criteria, it should be noted that the ways of determining whether the criteria have been met, differ between qualitative and quantitative research, and are expressed differently (hence the reference to ‘bilingual nomenclature’ earlier), but that should not detract from the equivalence between the two sets. There is a greater degree of standardisation of the criteria in the quantitative than the qualitative approaches. In the third place, meeting the criteria – in each methodological
approach – sets the limits of the extent to which the studies involved qualify as scientific, and it therefore becomes an imperative in each study to report the outcomes of determining or testing the reliability/dependability, internal validity/credibility, external validity/transferability and ‘objectivity’/confirmability of the study. (Even some quantitative survey studies omit to report the reliability and internal validity of the survey questionnaire used in the particular study, which implies conclusions such as, ‘We are 95% certain that we can generalise our findings to the relevant population – unfortunately we are not sure what we are generalising!’) In the fourth place, the fact that all the cells in Table 2 could be populated supports the argument offered earlier in this paper, namely that there is sufficient complementarity between quantitative and qualitative approaches to justify the argument that the utilisation of methodological approaches from both could contribute to a more nuanced approximation of the truth.

6.3 Root dimension 3: Primary purposes for which research is normally undertaken

Even a superficial reading of reports on research projects shows that a recurring characteristic that differentiates research projects is the goal of the project, or the so-called ‘teleological function’ of research. Lead researchers and senior academics classify research designs in a wide variety of ways (e.g. Blumberg, Cooper and Schindler, 2008: 196). These classifications are generally useful, but use different facets of research as the basis, such as research design, research setting, and information-collection techniques, to mention three. The Dutch social psychologist, Van Leent, as far back as 1965, offered a useful and, it is suggested, still relevant classification of the purpose of research that captures the teleological dimension, while allowing for linkages between quantitative and qualitative research. (Perhaps, this linking of quantitative and qualitative research should come as no surprise, since Van Leent wrote during the period of the flowering of phenomenology in northern Europe!).

Van Leent (1965) proposed a three-dimensional typology of the social sciences in general, and social psychology in particular. More specifically, he argued that the teleological function was a meaningful differentiator between various research traditions pursued in the social psychology of the day – and, we would argue, in the social sciences today - , and he identified the following basic goals of research:

6.3.1 Descriptive research

Many social scientific projects, including business and management studies, have the primary goal of describing a phenomenon (e.g. how many instances of the phenomenon exist in a population as found in a survey, such as marketing research). Normally, inductive theories (e.g. typologies) emerge from such research.

6.3.2 Explanatory research

A relatively small percentage of social scientific research tries to explain the exact nature of the relationship between phenomena (i.e. to determine cause–effect relationships), and an even smaller portion of those uses one of the traditional experimental designs to show the effect of the independent variable or cause on the dependent variable (Campbell and Stanley, 1963). In quantitative research, hypothetico-deductive theories normally emerge from this type of research design.

6.3.3 Understanding (German: ‘verstehende’) research

Describing the incidence or properties of a phenomenon (e.g. how many jobs have to be cut for a business to survive) and/or explaining the causal relationships among phenomena (e.g. the reasons for the need to reduce the workforce significantly) – important as they may be – does not tell us what it means to a loyal and hardworking employee to learn that one has been declared redundant. Qualitative research approaches, such as phenomenological, ethnographic and related designs and methods, can often contribute to in-depth understanding of the experience - retrenchment in this case. Descriptors of qualitative research (such as insight and understanding from the perspective and experience of the subject) bring this goal of research quite well to the fore.

6.3.4 Summary of teleological dimension

Normally, but not always, quantitative research approaches take the form of descriptive or explanatory studies, while qualitative studies are more focused on in-depth understanding of one or more phenomena. Van Leent’s (1965) original typology of social psychology – and, as this paper implies,
the larger part of research methodology in the social sciences in general – can be represented by the three axes of a cube (Figure 2).

Figure 2: Three fundamental goals of scientific research (adapted from Van Leent, 1965)

7. The framework: Research methodology in 3D

A three-dimensional framework emerges when the three root dimensions, namely constituent components, epistemic criteria and primary goals of research, are joined up, as shown in Figure 3.

Figure 3: Research methodology in 3D

Figure 3 reflects the following salient aspects:

- The framework can be considered as a decision-making tool for both planning and evaluating research.
- Each of the dimensions summarises the building blocks that constitute scientific research and simultaneously shows the equivalences between qualitative and quantitative research.
- Each of the dimensions can be – and normally is – used individually as specifications in planning, undertaking and/or assessing a project of limited scope.
However, the framework offers a ‘joined-up’ space that facilitates the planning of complex multi-, inter- and transdisciplinary research projects, especially if the boundaries between qualitative and quantitative approaches have to be crossed in the same venture. The framework highlights the elements that initially have to be identified and the criteria that have to be specified. Experience shows that an interdisciplinary academic research project would not follow exactly the same routes as a commissioned transdisciplinary research project, but both would have to account for each of the checkpoints in the cube (i.e. components and process criteria, epistemic criteria and research goal). Planning of such complex research projects requires the plotting of each of the research approaches to be utilised (quantitative and qualitative) on each of the three root dimensions to determine their compliance with the three sets of scientific requirements (components, criteria and research goals). A useful tool for this purpose would be a radar or spider chart, being a two-dimensional representation of multivariate information (e.g. Wikipedia, 2012); the use of such a conversion of the 3D Framework to a radar chart is shown in section 8.1.

The 3D Framework could also be useful in assessing all the elements included in the project design of a comprehensive multi-methodological research project.

Finally, the framework could be used as a set of planning specifications irrespective of whether the researcher’s orientation is qualitative or quantitative.

8. Potential contribution of the 3D Framework

The paper shows that the root dimensions constituting the framework comply with expositions of research methodology found in most standard textbooks on this vast area of study, but go beyond some by, firstly, explicitly relating the three root dimensions and their subsets to form an interlinked framework. Secondly, the paper assumes and demonstrates salient tangent planes between approaches to scientific research that have traditionally been assumed to be mutually exclusive, namely quantitative and qualitative research. Thirdly, the framework perhaps goes further than many traditional presentations and partially reinforcing more recent ones (e.g. Cresswell, 2003) by demonstrating the possible complementarity between research approaches. Finally, the framework further addresses latent imbalances that emerged from the modest survey of the papers presented at the ECRM 10 and 11 conferences (see section 1).

The framework, as it stands, has its limitations. In the first place, the author of the paper probably could not escape from his own socialization in quantitative methodological approaches. Secondly, the framework has been developed and utilised in a behavioural and social sciences context only and has not been subjected to a thorough cross-disciplinary peer review process. Thirdly, an inherent limitation of the framework is that it is selective to the extent that it incorporates what the author regarded as the most important and salient elements. Other scholars, depending on their own methodological preferences, might argue that other elements should have been incorporated in the framework. Fourthly, the relative complexity of the 3D Framework as represented in Figure 3 should be noted, but developing radar or spider charts would largely compensate for this factor. These and other limitations are receiving further attention.

The following section demonstrates the application of the framework by means of a radar chart.

8.1 Comparative evaluation of two projects

Two dissertations that were completed under the supervision of the author in 2000 were subjected to trial evaluations using the 3D Framework. These two projects formed part of a comprehensive research programme in a deep-rural part of South Africa, the Role of Education and Training in Rural Human Development (RETiRHD; Marais, 2001), which consisted of nine master’s projects – three qualitative studies, incl. one focus group-based project; three field experiments; one quantitative correlational study; one regional postal survey (n=500); one youth survey (34 villages, n=500); and one regional survey (random sample, n=1 000, covering an area of ±11 000km²). The two projects were selected because they represented a qualitative and a quantitative approach, respectively. The qualitative project (referred to here as ‘School Gov Skills’) was aimed at unpacking the role of rural school governing skills in the achievements of deep-rural schools and relied on focus group techniques for information gathering. The second project (referred to here as Small Business Skills) was designed to determine the effectiveness of training deep-rural people in small business skills over five days and used a quantitative correlational design.
For the purposes of the present demonstration the three root dimensions (i.e. constituent components, epistemic criteria and primary goal of the research) were allocated equal weights on the assumption that a wrong decision on any of the three or an incorrect application of the components of each would have negative effects on the ‘scientifiveness’ of the project. The components of each of the root dimensions were allocated the following weights:

- **Constituent components**: Each of the four components (see Figure 1) was allocated a maximum weight of 5.
- **Epistemic criteria**: Each of the five epistemic criteria (see Table 2) was given a maximum weight of 4.
- **Primary goal of the research**: The appropriateness of the goal, that is largely the design of the project (see Figure 2) and the correct application thereof were each given a weight of 10, totalling 20.

The sources of information upon which the specific scores for each of the two projects were based, were the original assessment of the supervisor (author of this paper) supported by information gleaned from the examiners’ reports. Figure 4 summarises the results of this application of the 3D Framework.

![Figure 4: Application (radar chart) of the 3D Framework to two projects](image)

This outcome of comparing two projects by means of a radar chart application of the 3D Framework is self-explanatory, but the following points are worth highlighting: Firstly, the performance of projects on the root dimensions of the 3D Framework can be plotted on a two-dimensional plane. Similarly, the extent to which the root dimensions are represented in a research proposal can be determined. Secondly, Figure 4 shows the similarities and differences between the two illustrative projects and this could serve as a useful quality control measure in the original project planning and subsequently in the research evaluation phase. Thirdly, the radar chart shows that the qualitative study met the epistemic criteria to a somewhat greater extent than did the quantitative study, while the latter was judged to be located closer to the process criteria of the constituent components as reflected in Figure 1. Fourthly, for illustrative purposes the root dimensions were given equal weights, but from a methodological perspective one could argue that differential weights would be more appropriate. Finally, it should be reiterated that this comparative analysis was undertaken by the author, who was the programme leader of RETIRHD, the programme of which these two projects formed part, the supervisor of the two master’s students whose dissertations have been used as examples, as well as having been one of their examiners. This section is consequently intended as an example of how the 3D Framework could be applied, rather than as an empirical test of that framework.

9. Functions of the framework

The root dimensions have served as a framework for the author as an active researcher and teacher for many years, but in a linear rather than integrated mode. This paper is the first attempt at linking the
three so-called root dimensions to form an integrated three-dimensional framework. The root dimensions, individually and in combination have proved to be facilitating mechanisms in teaching, project planning and evaluation.

9.1 Teaching of research methodology

The root dimensions and framework have proved useful as a map to assist teachers in systematising the logic of the research process at public universities (with relatively inexperienced students) and private universities (where students tend to come from professional occupations) and to aid students in charting their way through what some experience as a maze of new concepts and seemingly contradictory criteria (see Mkansi, Acheampong, Qi and Kondadi, 2012 for an analysis of some of these contradictions and the dilemmas students experience in this regard).

9.2 Project and research programme planning

The framework has proved useful as a planning instrument for complex research programmes – e.g. a programme matrix of 40 individual projects from which overarching integrated conclusions and recommendations were required (see HSRC, 1987: 3–8). Clearly, the root dimensions, individually and in combination, can serve equally useful project planning functions when critical questions on the scientific constituent components, scientific criteria and purpose of the project have to be considered.

9.3 Evaluation of research proposals and reports

The combination of root dimensions, as reflected in the Research Methodology in 3D framework, can serve as a useful assessment guide in evaluating research proposals and reports. In this regard, the framework represents a logical set of evaluative dimensions that can be checked, irrespective of the specific methodological approach of the researcher, especially but not exclusively in the evaluation phase of what Venable and Baskerville (2012) referred to as a design science of research methods. In practice, this means that each of the first two root dimensions could be transformed to a set of five and four summative rating scales, while the third root dimension would normally consist of two subscales measuring the appropriateness and the correctness of application of the chosen research goal, producing 11 rating scales in total. (The demonstration in section 8.1 summed the rating scales over each of the three root dimension.)

9.4 Other functions of the framework

A number of further functions can be listed, such as serving as a common framework for debates between proponents of different methodological orientations within the social sciences and between the social and natural sciences and as a framework for identifying apparent commonalities and differences between, for instance, scientific research and the creative arts.

10. Conclusions

None of the individual pieces of information offered in this paper should come as new knowledge to the experienced researcher and perceptive teacher of research methodology. The paper is an integration of existing knowledge, as the dates of some of the references should show! The paper may open three perspectives for some readers – in varying degrees for advanced research students, active researchers and teachers of research methodology. Firstly, broad agreement on the definition of scientific research and its essential properties could take the sting out of the main, and often unproductive, debate between proponents of quantitative and qualitative approaches to research. Secondly, most social scientific research (including business and management fields) is multi-faceted and requires multi-method designs, which would include quantitative and qualitative approaches. This would apply even more in the event of possible practical, economic, human and associated consequences related to the eventual implementation of the research findings. Thirdly, the emerging framework, Research Methodology in 3D, brings to the fore the interaction effect between the three seemingly independent root dimensions of research, thereby offering a conceptual instrument for relating quantitative and qualitative approaches to research.

The following conclusions seem justified: Firstly, typologies of research approaches (e.g. qualitative and quantitative research) are often the results of conceptual exercises rather than empirical analyses of actual published research reports and consequently do not necessarily represent research as practiced. Secondly, the framework offers a conceptual tool that could be used to find interfaces
between methodological approaches, while at the same time highlighting the unique contribution of each. Finally, perhaps the time has come to seek complementarity between approaches rather than spending time trying to defend one’s own approach and losing sight of the potential contribution of other approaches – even within one’s own paradigm.

Acknowledgements

Financial and other support of the University of Pretoria’s Graduate School of Technology Management is gratefully acknowledged. Tinus Pretorius (University of Pretoria) and Magdal Pienaar-Maraist (National Research Foundation) are thanked for their constructive comments on a draft of the paper. Marlene Mulder (University of Pretoria) is thanked for the design of some of the figures and the copy-editing done by Robyn Arnold deserves a special word of appreciation.

References


Looking at the Past to Enrich the Future: A Reflection on Klein and Myers’ Quality Criteria for Interpretive Research

Ana Cardoso and Isabel Ramos
Centre Algoritmi, University of Minho, Guimarães, Portugal
anachcardoso@dsi.uminho.pt
iramos@dsi.uminho.pt

Abstract: In the last two decades, interpretive research has become more established and more popular in information systems field. The work of Klein and Myers (1999) consists of a set of principles for conducting and evaluating interpretive research, which provide fair and appropriate criteria for assessing the validity and reliability of such studies and, given the number of citations, has had a significant impact in the interpretive research literature. Our article focuses on understanding how this set of principles has informed research articles published in two of the highest-ranked information systems journals and, specifically, questions if these principles have been translated into common practices when conducting interpretive research in the field of information systems and whether authors incorporate them explicitly when they communicate the results of their research. We reviewed articles published in Management Information Systems Quarterly and Information Systems Research, collected any explicit or implicit evidence of quality criteria that informed the research, and highlighted direct or indirect reference to Klein and Myers criteria. We summarize and compare our findings in a comprehensive table, and note that, apparently, the principle of hermeneutic circle and the principle of suspicion are the most explicitly discussed in this sample. Moreover, Klein and Myers’ set of principles seem to have had a greater influence in the papers published in the period from 2002 to 2006. This study provides a reflection about methodological rigor in interpretive research that, to our knowledge, had never been done. Thus, the findings here presented may be useful for junior researchers and doctorate level students to understand how validity and quality criteria are enacted in high-quality interpretive research and, we hope, may encourage them to build on the exemplary work of the authors we reviewed and thus to contribute to enriching the literature of qualitative research methodology in information systems field.

Keywords: interpretive research evaluation, quality and rigor criteria, information systems, Klein & Myers’ set of principles, hermeneutics, phenomenology.

1. Introduction

Research in information systems (IS) has had a long positivist tradition. However, qualitative research methods, in particular interpretive studies, have been gaining their own place within IS research community, spurred by serious and valuable contributions under various themes that have been published in past decades (Walsham 1995b) (Walsham 2006) (Trauth 2001). Publications with detailed information and practical guidelines for conducting this type of studies have also contributed to a growing confidence and to building a cumulative tradition in interpretive IS literature, as for example (Walsham 1995a), (Walsham 1995b), (Walsham 2006), (Myers 1997) (Klein & Myers 2001).

The problem of rigor and relevance of interpretive research is central in the debate about using appropriate research methods in the IS field (Mingers 2003) because interpretive researchers often face misjudgements about the rigor of their research, as they are sometimes required to satisfy criteria that are not adequate to their genre of interpretive research (Sarker 2007), (Klein & Myers 1999). The paper entitled “A Set of Principles for Conducting and Evaluating Interpretive Field Studies”, by Klein and Myers (1999) is a significant contribution to the topic of quality and validity criteria for interpretive research in IS field. Several pieces of interpretive studies have been published since. We collected a sample of high-quality interpretive research papers published after 1999 and assessed the methodological rigor and quality criteria concerns that have driven the researchers. In doing so, we highlight the explicit reference to Klein and Myers’ set of principles, the implicit form of the principles in the report, and any other quality and validity criteria cited by the authors.

This work aims at contributing to the literature of qualitative research methodology in IS area, in particular interpretive studies. The interested audience are junior researchers and doctorate level students that intend to undertake interpretive fieldwork. Our contribution provides an informative and educational account of how authors of these top-quality journals conceive methodological rigor when they communicate their research and, particularly, how Klein and Myers’ (K&M) set of principles are enacted in the conduct of interpretive filed studies. In terms of contribution for the IS field, we attempt at mapping how well-defined quality criteria have been applied in interpretive research and we provide an annotated bibliography of field studies developed under the tradition of interpretivism. In terms of...
contribution for qualitative research methods, we identify other quality and rigor criteria that have been used and that may be useful as a reference for researchers conducting this type of studies. Our contribution can help junior level researchers develop a critical view of their readings by having as a reference this sample of high quality articles while, at the same time, stimulate future interpretive studies to build on the best examples we have in the field and enrich the interpretive research literature.

The structure of this paper is as follows: the next section discusses quality and validity criteria for conducting and evaluating interpretive research in the context of IS literature. In section 3, we explain the methodology for this article, that is, how did we get to the results, and we discuss the results in section 4. The article terminates in section 5, where we draw conclusions and elaborate on limitations and future work.

2. Quality and validity criteria for interpretive research

Interpretivist research poses as epistemological assumption that knowledge about the world is acquired through social constructions such as language, consciousness, and shared meanings. The focus is on human sense making of the situations as they occur and on the meanings people assign to the situations. There are two strands of thought of interpretive research based on this epistemological assumption: one is based on language and its meaning; the other is more related to phenomenology and hermeneutics (Klein & Myers 2001).

The concerns about validity and quality of a research contribution are not exclusive of interpretive research. To establish credibility and validity of a research study for an audience is necessary to demonstrate care in reporting, which is both a matter or argumentation style and accuracy of methods used (Walsham 1995a). The justification of methodological approaches is often done on the basis of criteria utilized or developed by very experienced researchers in particular methods, disciplines or strands of thought. For example, the dimensions of authenticity, plausibility, and criticality (Golden-Biddle & Locke 1993) that ethnographic texts are required to develop in order to create a convincing argument.

Another example is the set of principles proposed by Klein and Myers (2009). These principles explore the role of hermeneutics, anthropology and phenomenology in eliciting knowledge from interpretive studies. Adherence to the principles does not guarantee interesting results per se but, instead, they are useful for orienting the conduct and report of interpretive research and “can also be used for post hoc evaluation” (Klein & Myers 1999, p. 71). However, they advert that researchers should not apply the principles mechanistically but reflect how and if they are useful to any particular project. The principles overarching the research process and provide tools for understanding the problem studied as a whole. Therefore, the resulting report will very likely be plausible and convincing.

The hermeneutic principle is a fundamental meta-principle, and the other six principles are in its dependency. The principles are as follows (in numerical order) (Klein & Myers 1999): the principle of hermeneutic circle; the principle of contextualization; the principle of interaction between the researcher and the subjects; the principle of abstraction and generalization; the principle of dialogical reasoning; the principle of multiple interpretations; and the principle of suspicion. These principles are not to be thought of in a bureaucratic way but instead in the spirit of fundamental ideas that embed insights from relevant strands of interpretivism, such as hermeneutics and phenomenology, and also from the authors’ experience of anthropological research.

Even though a thorough description of each principle can be found in the cited work, we next explain briefly each one. The fundamental meta-principle is the hermeneutic circle, which suggests that human understanding is achieved by iterating between the interdependent meaning of the parts and the whole they form. The parts and the whole can be at various levels, as for example different levels of observations (researcher and participants) and their shared understanding, various historical and social aspects and the whole perspective of the context, and also the partial understanding resulting from a study and the relation with the whole of the extant literature. The principle of contextualization thus follows from this idea, in that an explanation of the situation under investigation requires a critical understanding of the social and historical aspects.

The principle of interaction between the subjects and the researchers suggests that there should be a critical reflection of how research materials were socially constructed through the interaction between
researchers and participants. On the other hand, the principle of abstraction and generalization recommends that researchers relate the idiographic details revealed by the data interpretation to theoretical and more general level concepts. Furthermore, the principle of dialogical reasoning cautions that there should be sensitivity to possible contradictions between the theoretical framework guiding the research and the actual findings. The sixth principle underscores the sensitivity to possible differences in interpretations of the events under study among the participants, and the principle of suspicion follows from this one in that the researcher should also be sensible to possible biases and distortions in the narratives collected from the participants.

The methodological rigor of interpretive research is a concern that has lingered among qualitative IS researchers. Hence, the aforementioned set of principles was received with enthusiasm and, in many cases, was included in the mandatory readings for doctoral-level qualitative research courses. Therefore, it is reasonable to expect that future interpretive research would incorporate the principles as a reference work. This last assertion is what we wanted to confirm with our study and we report our findings in the next sections.

3. Methodology

The work of Klein and Myers (1999) has had a significant impact in IS research community and accounts for an impressive number of citations in, for example, Google Scholar: more than 2600 citations in November 2012. Thus, we wanted to learn about its influence in the two most reputed journals of IS research (according to the ranking of Association of Information Systems): Management Information Systems Quarterly (MISQ) and Information Systems Research (ISR). In the next section we explain how we sampled the articles we examined and in Section 3.2, we provide a detailed account of our study of these articles.

3.1 Bibliographical search

We limited the bibliographical search to a ten-year period (1999 - 2009) and to MISQ and ISR journals because they are representative of highest-quality research in the IS area. We acknowledge this limitation and contend that this work, being exploratory, may be extended in the future in order to include a larger basket of journals. Articles were collected from the database Business Source Premier, sourced by EBSCO, where a search was done with the keywords interpretive case stud* and interpretive ethnograph* in all text of MISQ and ISR journals.

We retrieved 104 articles from our search with these terms – 77 articles from MISQ and 27 articles from ISR – and then we applied a two-step filtering as follows. First, we perused all the articles and categorized those that were clearly quantitative, conceptual, or editorial articles. In a second step, we read thoroughly the methodological section of the remainder articles and categorized each one according to the authors’ description of the methodological approach. This categorization was based on the original terms authors used to describe the methods adopted in the article and was validated and discussed between the two authors of this paper. Our approach was similar to Mingers (2003) in that we strove to produce a small and consistent set of categories.

The results of the categorization are depicted in Figure 1 and Figure 2. We note that these percentages are relative only to the result of our bibliographical search with the aforementioned keywords and do not refer to the complete collection of articles published by MISQ and ISR during the ten-year period.

We found an interesting diversity of research approaches in the collection of 77 articles we collected from MISQ. While a significant number of articles are more on the conceptual or theoretical side or report quantitative research approaches, there is an interesting percentage of 34% of qualitative articles, including action research, critical research, interpretive case study, interpretive ethnography, positivistic case study, and other author-reported qualitative approaches (repertory grid and revealed causal mapping).

For the 27 articles from ISR, we also found a large number of conceptual / theoretical articles and quantitative research approaches. However, we could not find identical diversity in terms of qualitative approaches, with the 26% of such articles referring only to interpretive case study, interpretive ethnography, or positivistic case study.
Since the set of K&M' principles addresses specifically the quality standards of interpretive field studies, for this paper we selected only articles reporting this type of studies, that is ethnographic or case studies of interpretive nature. Thus, our bibliographical search resulted in a sample of 12 articles from MISQ and 6 articles from ISR reporting ethnographic or case studies of interpretive nature. Table 1 provides the details of each article and our comment on the quality and validity criteria evoked or implied from the report. For the sake of readability, articles’ details are displayed sequentially.

### 3.2 Examination of articles

Interpretive research articles are typically long and dense. In our case, the articles in the sample have an average of 25 pages length. We learned from the examples provided in Klein and Myers (2009) how to identify the principles in the narrative. Some articles are not explicit about any validation or quality criteria that informed their research and we tried to infer them in an implicit form, as Klein and Myers (2009) explain. Furthermore, we complemented our analysis with the operationalization of the principles as explained by O’hEocha et al. (2012), which is based on questions derived from
keywords in Klein and Myers’ article. In the following table, articles 1 to 6 are from ISR and the remainder 7 to 18 are from MISQ.

**Table 1: Details and methodological criteria of sampled articles**

<table>
<thead>
<tr>
<th>Article 1</th>
<th>(Choudhury &amp; Sabherwal 2003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Method</td>
<td>Multiple case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>1 diversified services company; 1 financial services firm; 1 software manufacturer; 1 large county in Southeast US; 1 large food product firm</td>
</tr>
<tr>
<td>Focus</td>
<td>Evolution of portfolios of control over the duration of outsourced information systems development (ISD) projects</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>New insights for the use of portfolios of control in the clients’ perspective</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>The report does not refer explicitly to Klein and Myers’ criteria. The analytical process is described as interpretive and an explanation of the data analysis process is provided together with the reflection of how the findings relate with the known literature, adding parallels, highlighting differences and drawing new questions.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Article 2</th>
<th>(Levina 2005)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Method</td>
<td>Ethnographic field study</td>
</tr>
<tr>
<td>Research Site</td>
<td>Web-based application development project</td>
</tr>
<tr>
<td>Focus</td>
<td>Understanding the actual collaborative practices on multiparty information systems development (ISD) projects, and how the diversity of professionals involved influences the system being designed</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Theoretical constructs to understand the cycle of multiparty collaboration on ISD projects</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>Strong in terms of theory building, generalization and reasoning. Multiple viewpoints are critically presented. Refers to an inductive, grounded fashion way of developing theory. Does not refer explicitly to Klein and Myers’ set of principles.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Article 3</th>
<th>(Sarker &amp; Sarker 2009)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Method</td>
<td>Case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>Multinational high tech organization</td>
</tr>
<tr>
<td>Focus</td>
<td>Development of an empirically grounded framework of agility in ISD settings and identification of some actionable tactics for enhancing agility within such settings</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Authors refer specifically to Klein and Myers’ set of principles in the methodological section and illustrate with examples from how they conducted the study. In the same section, they refer also other methodological considerations and cite other authors, namely (Bryant &amp; Charmaz 2007), (Charmaz 2000), (Flick 1998), (Myers &amp; Newman 2007), and (Patton 1990).</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Article 4</th>
<th>(Schultze &amp; Orlikowski 2004)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Method</td>
<td>Ethnographic field study</td>
</tr>
<tr>
<td>Research Site</td>
<td>A bricks-and-clicks dot-com that uses self-serve technology</td>
</tr>
<tr>
<td>Focus</td>
<td>Implications of inserting self-serve technology into embedded relationships at the inter-firm micro level</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Managing the complementarity between relationships and IT in practice is fraught with tension at the micro-level with negative consequences for business strategies</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>The intellectual basis of the research is clear and mindful of the findings being contradictory with some established theories. Moreover, the article examines alternative views and notes different values and opinions expressed by strategic and tactical teams. Authors do not refer explicitly Klein and Myers’ set of principles.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Article 5</th>
<th>(Vaast &amp; Walsham 2009)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Method</td>
<td>Case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>Public Health Administration of France</td>
</tr>
<tr>
<td>Focus</td>
<td>Practice-based learning dynamics that emerge in networks of practice</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Formulates a model of trans-situated learning supported by an information infrastructure</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>Authors acknowledge that the study was informed by the set of principles of Klein and Myers, and explain that the data analysis was made in consideration of the principle of abstraction and generalization, principle of hermeneutic circle and principle of suspicion. We could also find implicit evidence of the other 4 principles.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Article 6</td>
<td>(Vidgen &amp; Wang 2009)</td>
</tr>
<tr>
<td>Research Method</td>
<td>Multiple case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>Two software development teams</td>
</tr>
<tr>
<td>Focus</td>
<td>Study the factors that enable or inhibit agility in agile software development</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Framework for the organization of agile software development</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>There is no explicit reference to Klein and Myers’ set of principles. The study is accounted as interpretive and we could find implicit evidence of principle of abstraction and generalization and principle of multiple interpretations.</td>
</tr>
<tr>
<td>Article 7</td>
<td>(Backhouse, Hsu &amp; Silva 2006)</td>
</tr>
<tr>
<td>Research Method</td>
<td>Case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>British standard in information systems security</td>
</tr>
<tr>
<td>Focus</td>
<td>The influence of exogenous contingencies for the creation of a de jure standard and reflection about the power mechanisms required for a standard to evolve</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Deep understanding of how human, social and economical aspects influence the development of a standard and theorization about power circuits</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>In the research methodology section authors provide a table with their reflection of how they validate their interpretations in the light of Klein and Myers set of criteria.</td>
</tr>
<tr>
<td>Article 8</td>
<td>(Davidson 2002)</td>
</tr>
<tr>
<td>Research Method</td>
<td>Case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>Health care insurance company in Eastern United States</td>
</tr>
<tr>
<td>Focus</td>
<td>Requirements processes in organizations and the socio-cognitive basis of power in ISD</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Develops a socio-cognitive process model of how frames and shifts in frame influence sense-making during requirements identification on ISD projects</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>The researcher provides an appendix with criteria for assessing the quality and rigor of her work, where she directly refers to Klein and Myers’ set of principles and to criteria used in ethnographic texts, namely authenticity, plausibility, and criticality (Golden-Biddle &amp; Locke 1993)</td>
</tr>
<tr>
<td>Article 9</td>
<td>(Davidson &amp; Chismar 2007)</td>
</tr>
<tr>
<td>Research Method</td>
<td>Case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>Hospital in United States</td>
</tr>
<tr>
<td>Focus</td>
<td>Interaction between changes in social structures of an institution and changes in technology to engender alignment</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Uncovers two processes of institutionally-triggered changes and one process of technology triggered change related with the implementation of a computerized physician order system</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>The article elaborates on how the authors applied the principle of hermeneutic circle and refers to Klein and Myers’ set of principles. The other principles were not explicitly illustrated in the report, but we could find them in an implicit form.</td>
</tr>
<tr>
<td>Article 10</td>
<td>(Garud &amp; Kumaraswamy 2005)</td>
</tr>
<tr>
<td>Research Method</td>
<td>Naturalistic inquiry wherein insights are induced through interpretive means</td>
</tr>
<tr>
<td>Research Site</td>
<td>Infosys Technologies</td>
</tr>
<tr>
<td>Focus</td>
<td>Employs a system perspective as an interpretive frame to explore how Infosys manages knowledge processes at and across the individual, group, and organizational level</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>The organizational initiatives undertaken to harness existing knowledge systems in general generate a virtuous cycle of knowledge accumulation, reuse, and renewal but unintended consequences may arise because of the mutually causal knowledge processes unfolding at and across different organizational levels. Knowledge managers need to intervene to steer the knowledge system and address the interactions among knowledge process drivers.</td>
</tr>
<tr>
<td>Article 11</td>
<td>(Gopal &amp; Pushkala 2000)</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Research Method</td>
<td>Field study under the methodology of symbolic interaction</td>
</tr>
<tr>
<td>Research Site</td>
<td>Western Canadian School</td>
</tr>
<tr>
<td>Focus</td>
<td>Explore the alternative epistemological stance of symbolic interactionism to look at GDSS use phenomenon</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>In order to improve our understanding of the GDSS phenomenon, it is necessary to shift the focus from the technology itself to the symbolic context in which it is enacted, that is the arena of human interaction. Moreover, the multiple meanings arising from this context make it unsuitable for generalization because it leads to inconsistent results, and calls for investigation focusing on technology in action.</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>Authors do not refer explicitly to K&amp;M’ set of criteria or any other criteria. They acknowledge that grounded theory tenets informed their choice of methodology. The system perspective they introduced resemble, in terms of how they present the results, the principle hermeneutical circle.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Article 12</th>
<th>(Hanseth et al. 2006)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Method</td>
<td>Case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>Norwegian hospital</td>
</tr>
<tr>
<td>Focus</td>
<td>The intended and unintended consequences of the implementation of a standard for an electronic patient record</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Demonstrate the socio-technical complexity of IS standardization and the related disorders created by multiple actors that generates reflexive mechanisms and undermine the creation of a standard</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>Authors acknowledge that Klein and Myers’ set of principles informed data analysis and collection and illustrate with the fundamental principle of hermeneutic circle, though without naming it. We could find implicit evidence of the other principles.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Article 13</th>
<th>(Levina &amp; Vaast 2005)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Method</td>
<td>Multiple case study using grounded theory techniques</td>
</tr>
<tr>
<td>Research Site</td>
<td>Insurance company and an interorganizational IS development project involving an Internet consulting firm and its client</td>
</tr>
<tr>
<td>Focus</td>
<td>Improved understanding of the dynamics involving agents and artifacts becoming boundary spanners-in-practice and boundary objects-in-use.</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Investigation of how organizational competence in boundary spanning emerges in practice, which is related to the engagement of boundary spanning agents in a new joint field through which these agents discover an interest in spanning boundaries.</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>The article particularizes the principle of suspicion of K&amp;M’ set of principles. Another guiding principle explicitly acknowledged was radical doubt (Bourdieu &amp; Wacquant 1992), and authors also mention that they used grounded theory techniques to compare and contrast the independently-created case descriptions, both within and across cases. We found implicitly evidence of some other K&amp;M’ principles in the report.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Article 14</th>
<th>(Levina &amp; Vaast 2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Method</td>
<td>Draw on practice theory to analyse and interpret data from two empirical cases</td>
</tr>
<tr>
<td>Research Site</td>
<td>Large multinational, financial services firm headquartered in United States and Western Europe that outsourced projects to Russia and India</td>
</tr>
<tr>
<td>Focus</td>
<td>How differences in organizations and countries give rise to boundaries and status differences in offshore ISD projects and what practices alleviate these differences to establish multiparty collaboration</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Theory-based practice framework that explains how boundaries and status differences limit collaboration effectiveness in offshore ISD and the key role of on-shore middle managers in engaging others in shared practices</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>The article refers to Klein and Myers’ set of principles but does not explicitly particularize any principle. We could imply from the report that the principle of interaction between the researchers and the subjects was applied to improve inductive theory. Authors also refer that the principles of grounded theory informed data collection and analysis.</td>
</tr>
<tr>
<td>Article 15</td>
<td>(Olsson et al. 2008)</td>
</tr>
<tr>
<td>Research Method</td>
<td>Multiple case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>Irish sites of two large companies based in United States</td>
</tr>
<tr>
<td>Focus</td>
<td>Investigates and develops a theoretical model of the dual bridge role in a two-stage offshoring relationship</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Theoretical model that illustrates the off-shoring bridge model, where companies differ in terms of their approaches to team integration, organizational level implementation, and site hierarchy.</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>The principle of dialogical reasoning and the principle of multiple interpretations are explicitly evoked. Moreover, the principle of abstraction and generalization is clearly demonstrated in the description of the approach to data collection and analysis. The principle of interaction between the researchers and the subjects can be implied from the account of how the data was collected, and the principle of suspicion is demonstrated through quotes. Other criteria, besides Klein and Myers’ set of principles, are also explicitly justified and demonstrated namely reliability, construct validity and external validity.</td>
</tr>
<tr>
<td>Article 16</td>
<td>(Pawlowski &amp; Robey 2004)</td>
</tr>
<tr>
<td>Research Method</td>
<td>Case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>Large manufacturing and distribution company based in United States</td>
</tr>
<tr>
<td>Focus</td>
<td>The role of IT professionals in internal knowledge transfer, in particular as knowledge brokers</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Formulates a conceptual framework that shows the conditions, practices and consequences of knowledge brokering by IT professionals</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>We found explicit reference of Klein and Myers’ set of principles in the method section another section entitled “evaluation”, where the authors provide a reflection about their conformance to established principles for evaluating interpretive research. Besides demonstrating the strengths of their work in terms of adhering to some principles, as for example the hermeneutic circle, they also account for how the work could better reflect the principles of multiple interpretations and the principle of suspicion.</td>
</tr>
<tr>
<td>Article 17</td>
<td>(Puri 2007)</td>
</tr>
<tr>
<td>Research Method</td>
<td>Case study</td>
</tr>
<tr>
<td>Research Site</td>
<td>A district situated in the rain-shadow zone of peninsular India</td>
</tr>
<tr>
<td>Focus</td>
<td>The importance of knowledge alliances between multiple knowledge systems in order to support effective development and implementation of a GIS to address the problem of land degradation in India</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Provides an articulation of the concept of building knowledge alliances across communities of practice and broadens the concept of boundary objects</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>Author cites Klein and Myers’ set of principles but does not elaborate explicitly on any particular principle. We could find implicit evidence of the principles in the report.</td>
</tr>
<tr>
<td>Article 18</td>
<td>(Schultze 2000)</td>
</tr>
<tr>
<td>Research Method</td>
<td>Ethnographic field study with a confessional genre and a social constructionist stance</td>
</tr>
<tr>
<td>Research Site</td>
<td>Large Fortune 500 manufacturing firm headquartered in the Midwest of United States</td>
</tr>
<tr>
<td>Focus</td>
<td>Explore the work practices of knowledge workers</td>
</tr>
<tr>
<td>Main Contribution</td>
<td>Identifies three informing practices of knowledge work: ex-pressing, monitoring, and translating, and highlights the fundamental underlying logic of balancing subjectivity and objectivity; Further, the article demonstrates that these practices are trans-situational, meaning that they occur irrespective of “types” of knowledge workers.</td>
</tr>
<tr>
<td>Methodological criteria</td>
<td>We found no explicit reference to K&amp;M’ set of principles. The general evaluation criteria of ethnographic research (Golden-Biddle &amp; Locke 1993) – authenticity, plausibility, and criticality – is deemed as inappropriate for a confessional genre of ethnography and thus the author elaborates on the following evaluation criteria: authenticity; plausibility; criticality; self-revealing writing; interlacing actual and confessional content.</td>
</tr>
</tbody>
</table>
4. Discussion

In Table 2, we summarize our detailed assessment of quality and validity criteria in the articles. This process had two phases: in the first phase, we examined carefully all articles and registered our impressions about the methodological criteria reported by the authors for each article (see Table 1). We also registered, (in Table 2), explicitly evoked principles (X) and any other criteria (yes (Y) or no (N) when no other criteria was referred). In the second phase, we did a comparison among the articles in terms of the type of implicit support we could find for each principle: either strong implicit support (S), or weak implicit support (W) and registered the results of that comparison in Table 2.

Table 2: Summary of results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Choudhury</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>W</td>
<td>N</td>
</tr>
<tr>
<td>2. Levina</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>Y</td>
</tr>
<tr>
<td>3. Sarker &amp; Sarker</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>4. Schultze &amp; Orlowski</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td>5. Vaast &amp; Walsham</td>
<td>X</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>X</td>
</tr>
<tr>
<td>7. Backhouse</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>8. Davidson</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>9. Davidson &amp; Chismar</td>
<td>X</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>10. Garud &amp; Kumaraswa</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td>11. Gopal &amp; Prasad</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>Y</td>
</tr>
<tr>
<td>12. Hanseth et al</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td>13. Levina &amp; Vaast 05</td>
<td>S</td>
<td>S</td>
<td>W</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>14. Levina &amp; Vaast 08</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>Y</td>
</tr>
<tr>
<td>15. Olssom et al</td>
<td>S</td>
<td>W</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>16. Pawlosky &amp;</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>17. Puri</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td>18. Schultze</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>Y</td>
</tr>
</tbody>
</table>

Legend: W – weak implicit support; S – strong implicit support; X – explicitly evoked; Y – yes; N – no.

Notwithstanding the fact that the number of MISQ articles is superior in our sample, our results seem to show that Klein and Myers’ set of principles had a greater influence in the articles published in MISQ (art. 7 to 18), specially in the period shortly after the publishing of K&M’ article, where we could find several articles that explicitly state their conformation to that set of principles. These articles usually provide a highly structured explanation in the form of a comprehensive table or an annex where they justify their methodological considerations and provide examples for each principle.
Another finding is that the principle of hermeneutic circle and the principle of suspicion are, individually, the most cited and explicitly developed by the authors, probably due to the fundamental nature and overarching importance of hermeneutical process for interpretive research and to the significance of critical thought when building an integrated and coherent research report. On the other hand, it seems that, in most cases, when authors explicitly demonstrate the adherence to K&M’ set of principles; they choose to illustrate all the principles and not only a subset of these.

In terms of the implicit form of the principles, we found that the principles of abstraction and generalization and the principle of multiple interpretations can be implied in all articles. Indeed, the articles examined strive to generate a theoretical contribution that builds on the extant literature and to convey multiple accounts of the events under study. Conversely, the principle of contextualization is sometimes disregarded and minimized in the reports as some authors describe the case sites very broadly and do not provide a critical account of the historical context.

In terms of other quality and rigor criteria, we note that half of the papers we reviewed mention other criteria, which we believe is representative of diversity and richness of methodological literature within interpretive tradition. Particularly, we highlight article 11, because it lists criteria to evaluate symbolic interactionist studies, which are part of the phenomenological and social constructionism tradition of interpretivism (K&M’ principles also draw on phenomenology). We also note that four articles draw on grounded theory techniques (Glaser & Strauss 1967) and that seems to resonate with K&M’ principles because grounded theory incorporates techniques that are well suited to support the fundamental principle of hermeneutic circle, namely open and axial coding (Klein & Myers 2001).

Finally, a curious aspect is that four researchers authored more than one study in this sample of articles. In this respect, we could find some regularities in their reports. For example, articles number 2 and number 14 do not explicitly refer any K&M’ principle but acknowledge the adherence to grounded theory principles. Articles 8 and 9 are a different example. In this case, article number 8 explicitly refers to K&M’ set of principles, but the most recently published article (number 9) only explains how the hermeneutic circle was applied and does not delve into the other principles. Finally, articles 5, 13 and 14 also share one author but in this case we could not find a salient pattern in the report style.

5. Conclusion

Whereas the principles of Klein and Myers are based on best practices of interpretive research that have been developed for decades in other disciplines, the question remains as to why the two journals analyzed do not promote more compellingly their operationalization by publishing articles that make clear reference to those principles. This question is not answered in our article and lacks other kind of research that inquires the international research community about the real value assigned to the principles of Klein and Myers in their research.

Future research should perform the same search in other information systems journals, in order to identify the quality and rigor criteria applied not only in field studies but also in other methods of interpretive research. Another interesting avenue for future research is a visualization of the network of collaboration among researchers who are co-authors of articles that report interpretive research studies and refer to the principles of Klein and Myers in order to understand the area of influence of these authors within the international research community of information systems and how it has evolved. Moreover, there is a need for understanding more clearly how rigor is demonstrated in articles in which the principles of Klein and Myers are not explicitly reported, so as to perceive the way rigor has been encouraged in the information systems journals. Interviews with editors of journals on this topic may help to understand the efforts that have been made to publish interpretive research in high quality journals.

Finally, we argue that articles that make explicit reference and operationalize the quality and rigor criteria used are, in fact, providing a dual service to the information systems research community (and to interpretive research community at large): not only they contribute to build a cumulative tradition of interpretive studies but, and most importantly, they are also helping other researchers who want or need to learn about interpretive methodology by giving them finest exemplars of interpretive studies that they may emulate in order to enrich the future of interpretivism.
Acknowledgements

We acknowledge support of Fundação para a Ciência e Tecnologia via Bolsa de Doutoramento SFRH/BD/60838/2009. We also acknowledge support of Fundos Feder – COMPETE and Fundação para a Ciência e Tecnologia via FCOMP-01-0124-FEDER022674.

References


Patton, MQ 1990, Qualitative evaluation and research methods, Sage, Newbury Park.


Walsham, G 1995b, 'The emergence of interpretivism in IS research', *Information systems research*, vol. 6, no. 4, pp. 376–394.
A Proposal and Evaluation of a Design Method in Design Science Research

Francis Gacenga, Aileen Cater-Steel, Mark Toleman and Wui-Gee Tan
University of Southern Queensland, Toowoomba, Australia
Gacenga@usq.edu.au
Aileen.Cater-Steel@usq.edu.au
Mark.Toleman@usq.edu.au
Wui-Gee.Tan@usq.edu.au

Abstract: Information Systems (IS) design science literature offers a plethora of findings on various aspects, such as the general steps in design science, problem identification, objectives of solutions, and evaluation of the artefacts. However, there appears to be a dearth of guidance on the design of the artefact itself, that is, on design per se. Information Technology Service Management (ITSM) practitioners report challenges in measuring and reporting the performance of ITSM. This area is identified as significant and lacking in research. We are developing a framework for performance measurement of ITSM investments in organisations. The ITSM performance measurement framework (PMF) will provide a basis of standardisation and performance comparison for organisations implementing ITSM. The focus of this paper is the design methodology for the PMF framework. Our work considers literature from IS design science and disciplines outside IS. Previous IS researchers developed frameworks that guide the IS design research process but they do not provide details of the design process. We extend their work by narrowing the focus on the design step found in IS design science approaches giving a detailed treatment of the design step and delve further into the design literature broadly. Of course, design is a very creative endeavour and may not be solely process driven. Using a design process from outside IS design science advocating “designerly ways of knowing” we apply the integrative Matching Analysis Projection Synthesis approach. We integrate an IS design science research framework with a macro cycle of analysis-projection-synthesis and a micro cycle of research-analysis-synthesis-realisation to create an approach for designing the PMF. Outcomes of the evaluation are also described and discussed. The design process proposed was of limited success but may be useful for other IS researchers seeking guidance on how to design the artefact in an IS design science project.

Keywords: design science research, IT service management, performance measurement framework, mixed methods research, matching analysis projection synthesis approach

1. Introduction

In a research project to develop a performance measurement framework (PMF) for IT service management (ITSM) we employ a multi-paradigmatic and multi-method approach that includes design science research (DSR). Our review of Information Systems (IS) design science literature did not uncover specific advice on the design step that is required in most DSR frameworks. The objective of this paper is to highlight the silence on design in DSR and to propose a design approach for our particular project. As well as being the research project that motivated the need to identify more concrete and detailed guidance for the design step in DSR, the ITSM PMF serves as an example application of DSR and the design approach proposed.

ITSM is a customer-oriented approach used by IT practitioners to manage IT operations organised around IT services. ITIL® is the most widely used ITSM framework and is based on a library of books that offer “best practices” for ITSM. ITIL is a lifecycle-based, process-oriented framework that organizations can use to create, design, deliver and maintain customer-focused IT services.

This project addresses an area that has been identified as significant and lacking in research. ITSM practitioners report challenges in measuring and reporting the performance of ITSM. We are developing a framework that can be used for performance measurement of ITSM investments in organizations. The framework will provide a basis of standardization and performance comparison for organizations implementing ITSM and could be used to show a relationship between ITSM investment and benefits from this investment. Our project develops a PMF for ITSM, an IT service management tool. The research contributes to IS design theory by describing the “design decisions and design knowledge that are intended to be manifested or encapsulated in an artifact, method, process or system” (Gregor 2002). The research contributes to the existing ITSM body of knowledge by addressing the gap that currently exists in ITSM performance measurement (Lahtela et al. 2010), performance measures and methods.

ISSN 1477-7029
©Academic Publishing International Ltd
Reference this paper as: Gacenga, F, Cater-Steel, A, Toleman, M and Tan, W-G. “A Proposal and Evaluation of a Design Method in Design Science Research” The Electronic Journal of Business Research Methods Volume 10 Issue 2 2012 (pp 89-100), available online at www.ejbrm.com
The focus of this paper is on the design methodology for the ITSM PMF. The project uses a multi-paradigmatic and mixed-methods approach based on behavioural science and design science paradigms. The mixed-methods approach is suitable to the project as it addresses the theoretical challenge using science and, based on the empirical knowledge gained, develops a practical solution using design science.

We use the Information Systems Design Research (ISDR) approach as proposed in Peffers et al. (2008) as a foundation. An overview of the design research process model and behavioural science research methods are shown in Figure 1.

Figure 1: Initial design science method process model (Based on Peffers et al. 2008)

In stage one we begin with a systematic literature review followed by a survey. In stage two we conduct case studies and develop the ITSM PMF artefact. In stage three we evaluate the prototype PMF and enhance it through further design and development.

The paper is structured as follows. First the IS design science literature is reviewed. The gap in the literature is explained, followed by our proposal to overcome the gap/limitations. We describe how the proposed design approach is implemented, and evaluated. In the discussion section, we reflect on the outcomes based on use of the proposed approach. Finally, a summary and limitations are provided in the conclusion.

2. Review of IS design science literature

According to Hevner et al.’s (2004) article, “design science creates and evaluates IT artefacts intended to solve identified organizational problems”. The study by Gregor and Jones describes design science as a sub-strand of a collection of constructive research approaches with a common emphasis of the central role of the artefact (2007). Peffers et al. (2008) prescribe six processes for design science: identify problem, define objectives of a solution, design and development, demonstration, evaluation, and communication. A case for leveraging design theory to improve the transparency and rigor of design research is demonstrated by Piirainen and Briggs (2011) who integrate the framework in Hevner et al. (2004) and Peffers et al. (2008) as well as the design theory in Walls et al. (1992) with that offered in Gregor and Jones (2007). Patas and Goeken’s article suggests interplay between behavioural and design-oriented research can be improved and draws a distinction between empirical and theoretical knowledge as well as non-artefact-centric and artefact-centric knowledge (2011).

A review of the IS design literature provides guidance on how to organize IS design research. We found that literature was available on design science research steps (Carlsson 2006; Hevner and
Chatterjee 2010; Hevner et al. 2004; March and Smith 1995; Offermann et al. 2009; Peffers et al. 2008; Piirainen et al. 2010), problem identification (Wieringa 2010), objectives of a solution (Carlsson 2007), evaluation (Cleven et al. 2009; Pries-Heje et al. 2008), project management (vom Brocke and Lippe 2010) as well as evaluations of studies using design science (Samuel-Ojo et al. 2010). Design science studies providing the actual specifics of the design step are summarized in Table 1 which extends Peffers et al. (2008) and Offermann et al. (2009). In a number of design science studies the design and build steps are not clearly demarcated.

### Table 1: Design step in the IS design science literature

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Design &amp; Development Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cole, Purao, Rossi &amp; Sein (2005)</td>
<td>Build (model, instantiate)</td>
</tr>
<tr>
<td>Hevner, March, Park &amp; Ram (2004)</td>
<td>Iterative search process, artefact</td>
</tr>
<tr>
<td>March &amp; Smith (1995)</td>
<td>Build</td>
</tr>
<tr>
<td>Nunamaker, Chen &amp; Purdin (1990-1)</td>
<td>Understand the studied domain, application of relevant scientific and technical knowledge, creation of alternatives, and synthesis and evaluation of proposed alternative solutions</td>
</tr>
<tr>
<td>Takeda, Veerkamp, Tomiyama &amp; Yoshikawa (1990)</td>
<td>Suggestion, development</td>
</tr>
<tr>
<td>Vaishnavi &amp; Keuchler (2008; Vaishnavi and Kuechler 2009)</td>
<td>Suggestion, development</td>
</tr>
<tr>
<td>Walls, Widmeyer &amp; El Sawy (1992)</td>
<td>Design method, meta design</td>
</tr>
<tr>
<td>Offermann, Levina, Schonherr &amp; Bub (2009)</td>
<td>Design artefact, literature research</td>
</tr>
<tr>
<td>Peffers, Tuunanen, Rothenberger, &amp; Chatterjee (2008)</td>
<td>Design and development</td>
</tr>
<tr>
<td>Sein, Henfridsson, Purao, Rossi &amp; Lindgren (2011)</td>
<td>Building, intervention and evaluation</td>
</tr>
<tr>
<td>McLaren, Head, Yuan &amp; Chan (McLaren et al. 2011)</td>
<td>Design and evaluation</td>
</tr>
</tbody>
</table>

We considered the approaches summarized in Table 1 for the design step of our project but each lacks detailed guidance. The build step offered in Cole et al. (2005) includes the steps “model and instantiate” and advocates using software engineering principles such as effective tools and reuse.

Advice on the design step offered in Hevner et al. proposes an iterative search process that would result in the artefact. The article concedes that “given the wicked nature of many IS design problems it may not be possible to determine the relevant means, ends or laws” and suggest that a way out, “is to search for satisfactory solutions” (2004).

Emphasis on building the artefact is highlighted in March and Smith, though the article does not provide details on what “build” entails. The article states that “build refers to the construction of the artefact, demonstrating that such an artefact can be constructed” (1995).

Advice on systems design in Nunamaker and Chen (1990) states that “design involves the understanding of the studied domain, the application of relevant scientific and technical knowledge, the creation of various alternatives, and the synthesis and evaluation of proposed alternative solutions”. They did not give details on how to apply relevant scientific and technical knowledge in creating the alternatives or how to synthesize them.

The process model in Peffers et al. (2008) is synthesized from process elements found in seven studies taken from IS and other disciplines. In the model, design and development “includes determining the artefact’s desired functionality and its architecture and then creating the actual artefact” (Peffers et al. 2008). The article further suggests that moving from objectives to design and development requires “knowledge of theory that can be brought to bear in a solution” but does not explain how this can be done.

A computable design process model for CAD systems in Takeda et al. (1990) presents a descriptive model, a cognitive model and a computable model. The cognitive model proposes a five-step cycle that includes “suggestion” and “development”. Suggestion entails generating “key concepts that solve the problem” and development “to construct candidates for the problem from the key concepts using various types of design knowledge” (Takeda et al. 1990). The article further identifies object and action levels as distinct levels in the designer’s mental activity. The approach was adapted in Vaishnavi and Keuchler (2008) and we also find it useful as a conceptual guide though it does not address the practical design issues we faced.
An elaboration on IS design theory such as the systems development lifecycle (SDLC) with emphasis on building and testing IS theories is provided by Walls et al. (1992). Their design science research framework proposes design method and meta design in the design stage. However, it lacks detail on how to conduct these steps.

Based on the work in Peffers et al. (2008) a three-process framework entailing problem identification, solution design and evaluation is proposed in Offermann et al. (2009). The article considers Matching Analysis Projection Synthesis (MAPS) but concludes that “for IS design science, it seems that methodology hasn’t advanced as far as to propose different methods for each process step” (2009). The article then proposes, in its IS design research framework, a design step split into artefact design and literature search. Details are not provided on how to perform the artefact design.

Our work, like that in Peffers et al. (2008) and Offerman et al. (2009) considers literature from IS design science as well as outside IS design. The approaches in Peffers et al. and Offerman et al. develop frameworks that guide the IS design research process and do not delve into the details of the design process. The action design research method proposed in Sein et al. (2011) based on action research advocates evaluation performed “in authentic settings” using the design steps of building, intervention and evaluation. The article describes building the artifact and does not distinguish this from design.

To summarize, IS design science literature delivers a rigorous and systematic process on the general IS design research approach but offers little guidance on the design and development step at the heart of creation of the artefact. Design is a creative endeavour and may not be solely process driven. Few articles in IS design science describe the design step in detail or give much guidance on design per se perhaps because design needs much ‘right brain’ as well as ‘left brain’ activity (Tovey 1984) and tends to be context specific. However there are design principles that have been espoused over the years.

3. Design gap in IS design science

Design is defined as “a goal-directed thinking process by which problems are analysed, objectives are defined and adjusted, proposals for solutions analysed, objectives are developed and the quality of those solutions is assessed” (Roozenburg and Eekels 1995). The gap in IS design science is identified by Offermann et al. who state, “artefact design is a creative engineering process. Not much guidance is provided in IS literature” (2009). The gap is described as the lack of design foundations in the axiomatic statements of the formal sciences, the empirical approaches of the natural sciences and the hermeneutic techniques of the humanities (Jonas 2007). The subject and object of design science is design and using the definition of Roozenburg and Eekels, design is “to conceive the idea for some artefact or system and/or to express the idea in an embodiable form” (1995).

Design science lays emphasis on systematic, testable and communicable methods. However it is instructive to reflect on the differences between design science and the science of design and engineering as highlighted in Cross (2002). In our reflections we grappled with the question of whether design needs to be scientific and whether design science advocates that design is scientific. These questions are captured by Cross (2002) who observed “a desire to ‘scientise’ design can be traced back to the 20th-Century Modern Movement in design”. There were aspirations to produce works of art and design based on scientific methods of objectivity and rationality. These aspirations to scientise design surfaced strongly again in the ‘design methods movement’ of the 1960s.

We heed the advice in Osterle et al. on the design step, that “artefacts should be created through generally accepted methods, be justified as much as possible and be contrasted with solutions already known in science and business” (2010).

4. Our approach to the design problem

Designer-researchers are viewed as a possible solution to the design problem (Cross 2002). Designer researchers combine scientific methods with “designerly ways of knowing” (Cross 1982). The article identifies five aspects of designerly ways of knowing: “Designers tackle ill-defined problems. Their mode of problem-solving is solution-focused. Their mode of thinking is ‘constructive’. They use ‘codes’ that translate abstract requirements into concrete objects. They use these codes to ‘read’ and ‘write’ in ‘object languages’” (Cross 1982).
We reviewed literature on the design step outside IS design science (Archer 1984; Chow and Jonas 2008; Cross 2002; Eekels and Roozenburg 1991; Roozenburg and Eekels 1995) and identified detailed design processes that would be useful for IS design science. There is a large body of design literature and we refer to the works by Bayazit (2004) and Cross (1993) for a summary review of forty years of design research which describe research from the early roots: De Stijil, the Bauhaus; to first generation design methods: Horst Rittel, Morris Asimov, L. Bruce Archer, Christopher Alexander on patterns, and Herbert A Simon; to second generation design methods; and finally to scientific research in design.

According to the research in Chow and Jonas (2008), after the first generation methods were rejected, the postmodernist attitude of no methods was followed by strong adoption of scientific methods for design research. The article argues that the current situation regarding methodology in design research is characterized by unproductive dualisms and proposes a generic process model, Matching Analysis Projection Synthesis (MAPS). We also considered an alternative design cycle and process proposed in Archer (1984) and a similar one in Roozenburg and Eekels (1995) which entails analysis, synthesis, simulation, and evaluation conducted in iterations of specification, design, properties and comparison.

We opted to adopt MAPS as the design guide for the design step in our project as MAPS is similar to the concepts of ‘the true’, ‘the ideal’ and ‘the real’ matching realism which is the underlying philosophy of the project. MAPS also presents an integrative design research medium as well as flexibility.

5. Designing a performance measurement framework for ITSM

In Table 2 we outline how we conducted the ITSM PMF project by applying behavioural science methods with an IS design science framework that integrates ‘designerly ways of knowing’ in the design step. This fusion of science and design is achieved in three stages and we outline the objective, method, paradigm, parent discipline, philosophy, data analysis, purpose, study phenomena, outcomes and contributions.

<table>
<thead>
<tr>
<th>Table 2: Fusion of science and design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Research/Design</strong></td>
</tr>
<tr>
<td><strong>Objective</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Method</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Paradigm</strong></td>
</tr>
<tr>
<td><strong>Parent Discipline</strong></td>
</tr>
<tr>
<td><strong>Philosophy</strong></td>
</tr>
<tr>
<td><strong>Data Analysis</strong></td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
</tr>
<tr>
<td><strong>Phenomena</strong></td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
</tr>
<tr>
<td><strong>Contribution</strong></td>
</tr>
</tbody>
</table>

The macro and micro cycles of MAPS are depicted in Figure 2 and Figure 3 respectively. In the macro cycle, analysis refers to ‘the true’ how it is today; projection is ‘the ideal’ how it should be; and synthesis is ‘the real’ how it is tomorrow (Jonas 2007).
Jonas (2007) advises that a hyper cyclic generic design process model results from combining the domains of knowing in the macro model (analysis, projection, synthesis) with the learning phases in the micro model (research, analysis, synthesis, realisation) as depicted in Figure 4.

5.1 Macro cycle of design process

The first step in the macro cycle of the design process, analysis, is undertaken using the findings of the initial stages of our project and provides us with information of ‘the true’, that is, how ITSM performance measurement is currently conducted in industry. Integrating the findings of the literature review, survey and case study we gained empirical knowledge as well as in-depth qualitative knowledge of the state of the art in ITSM performance measurement.

The second step in the macro cycle, projection, deals with ‘the ideal’ and through the literature review and content analysis of the case studies we get an understanding of how performance measurement could be in the ideal. We generate an initial model of the ITSM PMF as it should be, depicted in Figure 5.
Figure 4: Generic, hyper cyclic model of the design process (Source: Jonas 2007)

Figure 5: Model to measure the performance of ITSM

The model proposed in Figure 5 can be used by organizations to measure the performance of their ITSM as well as evaluate the metrics they are currently using for completeness in service orientation, financial and non-financial perspectives in broad economic terms. The metrics within each perspective are categorized into service, function, process and technology dimensions. These represent ITSM function, ITSM process and ITSM technology metrics while the service dimension includes metrics to account for end-to-end process outcomes. At the industry level, wide adoption of the performance measurement model would lead to standardization of ITSM performance measurement and enhance the ability of organizations implementing ITSM to benchmark against other organizations using like terms.

Synthesis, the third step in the macro cycle of the design process is achieved in our project by working with a panel of industry and academic experts to identify how to contextualize the ITSM PMF.
In this step the framework is compared against the performance measurement practices currently in place in organizations. The panel of experts contributes to the development of a method to evaluate the ITSM PMF.

5.2 Micro cycle of design process

In the micro cycle of our project, research involves the data collection through literature review, survey questionnaires and case studies. Analysis entails using the findings of the qualitative and quantitative data analysis from the project stage two. Synthesis involves employing the five aspects of ‘designerly ways of knowing’ described in Section 4:

- The problem was ill-defined as a PMF had not been previously developed and the dimensions of the framework were unknown;
- The focus was to develop a solution to the problem: ITSM practitioners experience challenges in measuring and reporting the performance of ITSM;
- A constructive mode of thinking was adopted to develop, populate and implement the PMF;
- Figure 5 is an example of the visual representation of codes used to translate the abstract requirements of dimensions to arrange metrics for the metrics catalogue;
- This model informed the database design (‘object language’) of the repository to store the elements of the PMF including metrics, benefits, processes, BSC perspective, challenges, and business sector.

Realization includes the evaluation and communication of the prototype.

For each phase of the macro cycle (analysis, projection and synthesis), four steps in the micro cycle were undertaken (research, analysis, synthesis and realization). Table 3 presents a summary showing the specific design elements for the macro and micro design cycles. We developed a prototype and stopped at proof of concept with further development possible in future projects. Several iterations on the design prototype were undertaken.

Table 3: Summary of hyper cyclic model applied to the ITSM PMF project

<table>
<thead>
<tr>
<th>Macro Cycle</th>
<th>Micro Cycle</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Research</td>
<td>Analysis</td>
<td>Synthesis</td>
<td>Realization</td>
</tr>
<tr>
<td>Analysis</td>
<td>Literature review findings</td>
<td>Systematic literature review</td>
<td>Catalogue of metrics</td>
<td>Communication: journal, conference and seminar papers</td>
</tr>
<tr>
<td>Projection</td>
<td>Survey and case study data</td>
<td>Quantitative analysis and content analysis</td>
<td>Cross case analysis</td>
<td>Demonstration of prototype</td>
</tr>
<tr>
<td>Synthesis</td>
<td>Interactions with panel of experts and ITSM practitioners</td>
<td>Content analysis</td>
<td>Performance measurement framework</td>
<td>Tested prototype, thesis</td>
</tr>
</tbody>
</table>

An ITSM metrics catalogue was developed to populate the ITSM PMF. Proof of concept involved demonstration of the use of the ITSM PMF. The demonstration aimed to solve the problem of measuring the performance of ITSM in the case study organizations. This involved using the ITSM PMF as a guide to selecting relevant and contextualized ITSM performance metrics for the organization and then generating meaningful ITSM performance reports for the business. The revised design science method process model is shown in Figure 6.

6. Evaluation of artefact

Evaluation is an important requirement in design science research (Cleven et al. 2009; Gregor and Jones 2007). To evaluate the ITSM PMF artefact, the study applies the guidance and fitness-utility model for design science research proposed in Gill and Hevner (2011). Evaluation of the developed artefact aims to answer the question “How well does it work?” (March and Smith 1995). In this section, the ITSM PMF artefact is evaluated for its usefulness and fitness, on its capability to assist the organisation to select contextualized metrics, generate relevant reports to the organisation and provide an integrated set of ITSM performance metrics.
Two public sector organisations provided the opportunity to perform the evaluation of the ITSM PMF and the results of the evaluation are used in the synthesis step of the macro cycle of the design process. To maintain confidentiality, these organisations are referred to as Case A and Case B.

Case A sponsored the project with financial and in-kind contributions. Two meetings were held with the Case A staff. The panel of ITSM practitioners included the Senior Director of Service Management, the Director of Strategy, Governance and Architecture, the Assistant Director of Governance, the Director of Service Planning and Performance, and a Principal Process Improvement Officer. Service performance reporting was a priority for Case A though it had not yet been implemented. Case A collected process metrics that satisfied their operational, and to a lesser extent, tactical level of management. The ITSM dashboard in use tracks correlation between what has changed at the business level and its impact on performance at the operational level. Case A had identified a need for service metrics presentable to the strategic level reporting to the organisation's Chief Executive Officer (CEO). An earlier attempt by managers to present an operational and tactical ITSM report at Case A's strategic level failed. The CEO and strategic level organisation staff rejected the operational report by the IS unit as irrelevant to their needs. The strategic level was not interested in ITSM process metrics. The ITSM managers were of the view that a lot of effort was required internally to generate ITSM reporting. In the past, the BSC had been used for service performance reporting, but that had tapered off over time with a number of personnel changes in the role of CIO. Now Case A's newly appointed CIO has rejuvenated interest in service reporting and has advertised an ITSM service reporting position. Case A is also reinstating an ITSM governance committee that is developing a service improvement road map with a process-maturity monitoring program. The organisation has an extensive list of ITSM process performance metrics with limited reporting. The metrics are mainly tool sourced. The managers sought to find reporting along the BSC perspectives with five to six levels of breakdown to allow reporting to different performance reporting audiences such as strategic, tactical, and operational with the key performance indicators linking into reports. The managers believed they needed a multidimensional framework with a stakeholder’s view. The stakeholders are perceived to be the end customer of metrics. Case A had been using internal resources to develop ITSM performance reporting and had not received assistance from external consultants.

Currently the ITSM performance metrics evolve from qualitative metrics, and as maturity is achieved they tend to include quantitative metrics. Case A placed more value on an ITSM PMF and less on an ITSM metrics catalogue. The managers expressed a need for a performance measurement framework that shows how to ‘roll up’ the metrics across the various organisation levels. Higher value
is placed on aligning the ITSM with the organisational performance and less value in ITSM performance measurement at the operational level. The type of reporting generated and how the reports are used drives the validity of ITSM performance measurement. The level of granularity changes with the needs and priorities at different organisation levels, over time and with senior management's “appetite” for ITSM. Case A preferred not to have too many categories for performance reporting in the ITSM performance measurement framework. A recurring need expressed by panel members was the need to show how the ITSM metrics at the operational level roll up to dashboards at the business reporting level. Given that different ITSM report customers are interested in different measures at different levels, varying granularity will be required and the ITSM performance measurement framework would need to be dynamic. Knowledge of IT service issues in the customer domain was increasing and there was recognition of the importance of customers.

An ITSM performance measurement challenge experienced by Case A was the issue of ITSM staff meeting response times but reporting at a later date, thereby skewing the metrics on call response times. Case A sought an ITSM performance measurement model adaptable to the organisation offering an approach to implementing metrics with advice on where to begin. The performance measurement model ideally should provide advice on whether to start at the top level or operational level of the organisation or whether to measure the overall organisation performance or unit level performance. At the operational level the ITSM performance measurement framework needs to be comprehensive to allow "slice and dice". Managers were concerned about demonstrating how the ITSM performance measurement supported the business.

The senior managers at Case A had not reviewed the ITSM performance measurement framework documents emailed to them earlier and made comments and provided feedback on the presentation on the basis of their current organisational needs. Case A was undergoing an overhaul in their performance reporting and had created a new structure for strategy, planning, governance and architecture that subsumed the previous benefits realisation reporting function of the information function of Case A. The information function senior manager advised that ITSM performance measurement represented a small fraction of the Case A's information function needs and direction.

As a consequence of the organisation's focus and needs, Case A declined to commit further resources and time to review the ITSM PMF, to verify its usefulness, or provide further feedback on improvements.

At Case B, the researcher presented the three components of the ITSM performance metrics framework to the Manager (Service Reporting). The feedback from Case B was that the PMF appeared to be valuable and that it addresses areas that would benefit the organisation. The Manager (Service Reporting) confirmed that the organisation already has an ITSM metrics catalogue, and that the effort required to populate the proposed ITSM catalogue with the elements appears to be 'enormous'. The Manager (Service Reporting) is interested in applying current metrics to work through an example based on the PMF.

7. Discussion

Upon reflection, the outcomes of the PMF project were not entirely satisfactory: the prototype PMF has not developed to the point of an actual industry trial. Temporal and financial constraints prevented further iterations through the design-build-evaluate cycle.

Engagement with the project sponsor (Case A) was stymied in part by organisational turbulence. Case A suffered a highly publicised IT system failure that resulted in high levels of uncertainty and staff turnover in the roles of the ITSM PMF project partners as well as the CIO. The evaluation by Case B occurred too late in the project timeframe to action Case B's comments.

We recognise that the effort involved in searching for a suitable design framework was not included in the time allocated for the project. Unfortunately, although the MAPS model satisfied the requirement for concrete and detailed guidance for the design step in this design science project, it may have prevented creativity and flexibility by imposing a too-rigid structure. Drawing from the Information Systems field, a more agile approach (Highsmith and Cockburn 2001) to the design and development of the PMF may have resulted in more frequent prototypes, earlier feedback from the practitioners and the ability to tailor the artefact to the requirements of the sponsor organisation.
8. Conclusions

In summary, the review of the IS design science literature revealed a scarcity of guidance on the design of the artefact, despite a substantial research effort on other aspects of the field, such as the general steps in design science, problem identification, objectives of solutions, evaluation of the artefacts and project management. Our project aimed to develop and evaluate a framework artefact that can be used for the performance measurement of ITSM investments in organizations. The project uses the Information Systems Design Research (ISDR) approach (Peffers et al. 2008). The objective of this paper has been achieved: the silence on design in IS design science has been highlighted and a design approach has been proposed and evaluated in our project.

We approached the fusion of science and design following the advice from the literature. After exploring the design gap in IS design science, we identified the MAPS approach from a discipline outside of IS and used it in our project. Using a design process from the design literature advocating 'designerly ways of knowing' we applied the integrative MAPS approach (Chow and Jonas 2008). We integrated the IS design science research approach with the MAPS model. Based on MAPS, we applied the macro cycle of analysis-projection-synthesis and the micro cycle of research-analysis-synthesis-realization to create an approach to design the prototype of the ITSM PMF.

The approach did not effectively guide the design to a satisfactory outcome. Our experience may shed light on why the design gap in design science is difficult to fill. We sound a warning to researchers that design is a creative endeavour. The design step can be supported, for example by design patterns, exemplars, past experience and approaches such as those from systems analysis. A concrete and rigorous design approach may stifle creativity, reduce agility and prevent flexibility. In our PMF project, a more agile approach with shorter cycles may have enabled more frequent engagement with practitioners and a framework more suited to their needs.

We recognize as a limitation to this work that there may be other design approaches from other disciplines, however, the design process proposed here provides an example and raises issues for other IS researchers faced with the dilemma of how to design the artefact in an IS design science project. This multi-discipline approach improved the rigor of our ITSM project by using a systematic and repeatable approach to guide the design step in developing the framework. This paper contributes to the body of knowledge on design science by proposing and evaluating a design approach to help fill the current gap.

Note: ITIL® is a registered trademark of the UK, Office of Government Commerce.

References


A Technical Guide to Effective and Accessible Web Surveys

Greg Baatard
Edith Cowan University, Perth, Australia
g.baatard@ecu.edu.au

Abstract: The Internet is becoming an increasingly prominent medium for the administration of surveys. Although individual findings vary, the majority of the literature agrees that the appropriateness and response rates of web surveys is expected to rise in the future as we enter a generation of “digital natives” and mail-based communication becomes increasingly antiquated. Just about every aspect and tenet of traditional survey methodology has received attention in academic literature, positioning it as one of the most mature data collection techniques and a mainstay in all areas of research. While much of this accumulated knowledge is applicable and relevant to web surveys, there are numerous issues that arise specifically when surveys are delivered online. Such issues concern the overall design, delivery and administration of web surveys and the accessibility, structure, presentation and layout of their questions. The decisions made in these areas can influence the efficacy of a web survey in a number of ways, including the rate, integrity and quality of responses. This paper discusses such issues, and makes a number of recommendations to assist researchers in manually developing an effective and accessible web survey and in evaluating survey creation products and services.

Keywords: web, online, survey, questionnaire, guide, accessibility

1. Introduction

Dillman, Smyth and Christian (2009) describe the past two decades as “turbulent times” for the survey methodology. The telephone-based surveys popular throughout the 1970s and 1980s have become substantially less popular amongst researchers, marred by numerous factors including a surge in telemarketing and the distasteful phenomenon of receiving calls during the “dinner hour”. While still heavily used and indeed the most suitable medium in some demographics, mail-based surveys have begun to lose ground with advances in technology spurring web surveys to prominence. As Internet access and eMail become almost synonymous with the ownership of or access to a personal computer, delivering self-administered surveys via such technologies has become an increasingly common practice. Doing so offers the same benefits of mail surveys – allowing the researcher to reach potential respondents all over the world with minimal cost of data collection and processing. Furthermore, web surveys require less effort by respondents to complete and return than mail surveys (Couper 2000; Fricker and Schonlau 2002; Lyons, Cude et al. 2005; Deutskens, de Ruyter et al. 2006; Couper and Miller 2008; Malhotra 2008; Dillman, Smyth et al. 2009; Vicente and Reis 2010). The medium also affords greater control in terms of validation and the delivery of the survey items; for example, conditional questions can be presented only if required (Oppenheim 1992; Couper 2000; Fricker and Schonlau 2002; Shropshire, Hawdon et al. 2009; Vicente and Reis 2010). The potential benefits of web surveys have been recognised for over a decade – long before the Internet rose to the prominent position it currently holds. Schmidt (1997) lists access to a large population, saving time and money and allowing for dynamic/interactive surveys as the benefits of publishing a survey on the Web. Potential problems are also listed: incomplete or unacceptable responses, multiple submissions, and the security and integrity of data. Fifteen years later, these issues have largely been resolved and the benefits can be readily observed.

The already low response rates typically seen in mail surveys are set to become lower still as more and more communication occurs electronically, making the completion and return of a printed survey a somewhat archaic process. Dillman, et al. (2009) echo this sentiment, stating that “the shift toward eMail as the communication mode of choice for significant sectors of the population is somewhat ironic as it is one of the very factors that make Internet surveys possible, but it is also making surveys by traditional modes more difficult to complete.” Web surveys normally require no more than eMail and/or a Web browser to complete and submit – both of which the large majority of today’s computer users are familiar with. Furthermore, the completion of trivial online quizzes, polls and tests has emerged as a social pastime for many Internet users, particularly those in teen and young adult demographics (Fricker and Schonlau 2002). Although done for amusement, these activities serve to familiarise Internet users with the mechanics of web surveys and potentially increase their willingness to respond to other surveys.

Several studies conducted in the past decade have found that web surveys can achieve similar response rates to mail surveys, particularly for younger respondents and those in demographics that
regularly use the Internet. Some of the studies finding response rates to mail surveys to be higher than those of web surveys have also acknowledged the need to take demographics into account and the likelihood of response rates differing in the future. These trends suggest that web surveys will become increasingly prominent and result in higher response rates as the population becomes increasingly made up of “digital natives” (Jones and Pitt 1999; Cook, Heath et al. 2000; Couper 2000; Couper, Traugott et al. 2001; Fricker and Schonlau 2002; Kaplowitz, Hadlock et al. 2004; van Selm and Jankowski 2006; Lusk, Delclos et al. 2007; Converse, Wolfe et al. 2008; Couper 2008; Couper and Miller 2008; Shih and Fan 2008). Indeed, the emergence of the Internet as a major medium for the delivery of self-administered surveys is easily observable and cannot be denied.

The issues of survey design are well established and mature in academic literature, covering topics such as question wording, question types and survey length. While such issues are entirely applicable and must be respected in web surveys, this paper focuses upon issues which arise specifically when surveys are delivered online. The issues discussed can have a significant impact upon the accessibility and usability of web surveys, potentially influencing both response rates and the quality of responses themselves – “Before writing a questionnaire for the web, it is important to remember that a poorly designed survey can discourage people from responding, and it can also give skewed results” (Gonzalez-Bañales and Adam 2007). Web accessibility is an increasingly prominent topic, with a number of governments recommending or requiring Web pages to meet accessibility standards. There are a number of considerations that should be made to ensure a web survey is accessible to those with impairments or disabilities. Several of the issues and recommendations are of a technical nature, written with the assumption that interested readers will have a moderate level of technical knowledge or at least an understanding of the technical aspects of developing and implementing a web survey.

2. Overall design, delivery and administration issues

This section discusses the primary issues related to the overall design, delivery and administration of web surveys as a whole. The first issue is that of the delivery and hosting of the survey. The author recommends against utilising eMail for anything other than contacting potential respondents, despite possibly appearing as a method of minimising the effort required to respond. While it is possible to embed or attach a survey to an eMail, this practice is relatively uncommon and is less likely to be supported in a uniform fashion by eMail client software (Fricker and Schonlau 2002; Lyons, Cude et al. 2005; List and Gosling 2007). Thus, response rates of eMail-based surveys may be hampered by both unfamiliarity and technical hurdles. An eMail used to contact potential respondents should be concise. It should introduce the survey, tell the reader how long it should take to complete, mention an incentive if one is offered, and present a clear link to the survey itself. eMail can also be used to follow up with respondents – for example to thank them for their response, remind them to complete the survey or advise them of any outcomes. Unless the survey is entirely anonymous, the researcher has the ability to ensure that eMails are not sent unnecessarily – do not disturb someone who has already responded with an eMail reminding them to complete the survey (Kaczmirek 2005; Gonzalez-Bañales and Adam 2007).

Web surveys should be hosted on the Internet as standard Web pages. They should minimise any reliance on supporting software or technologies such as Flash or JavaScript (Kaczmirek 2005; Gonzalez-Bañales and Adam 2007). As discussed in the next section, if such technologies are used they should not be relied upon and the survey should function properly without their presence. Such technologies also hinder the accessibility of a web survey, hindering its completion by those with some form of impairment or disability. Like all Web pages, all web surveys should be tested thoroughly before deployment to ensure that they appear and function correctly in all likely browsers, browser versions, screen resolutions, and so forth (Fricker and Schonlau 2002; Thompson, Surface et al. 2003; Kaczmirek 2005; Lyons, Cude et al. 2005; Gonzalez-Bañales and Adam 2007). The survey should be hosted on a Web site or domain that can be recognised as trustworthy and legitimate, such as that of an academic institution, a Web site related to the field of the research, or a well-recognised survey hosting Web site. Avoid hosting the survey on unrelated or advertising supported domains, as these can be perceived as unprofessional and untrustworthy (Kaczmirek 2005; Gonzalez-Bañales and Adam 2007).

Access to and availability of web surveys must also be considered. An all too common pitfall when conducting a web survey is to simply aim for the largest number of respondents possible – an attractive option when the costs of distributing, administering and collecting the survey have been
minimised by the medium. However, “large samples are no substitute for accurate samples” (List and Goslino 2003) – a medium that makes it easier to reach more people is no excuse to ignore appropriate sampling procedures (Simsek and Veiga 2001; Lyons, Cude et al. 2005; van Selm and Jankowski 2006; Couper and Miller 2008; Vicente and Reis 2010). In addition to adhering to appropriate sampling procedures, web surveys should require some form of unique identifier such as an ID number, code, IP address or eMail address to deter automated responses or multiple responses by a single person (Couper, Traugott et al. 2001; Thompson, Surface et al. 2003; Lyons, Cude et al. 2005). Such measures can increase the quality of the data gathered by deterring behaviour which negatively influences response data. When using an ID number or access code, Couper, et al. (2001) recommend avoiding those containing potentially ambiguous characters such as the letter “I” and the number “1” or the letter “o” and the number “0”. Limiting the availability of a web survey has the potential to strengthen the data set. For example, if a survey is to be administered only to employees of a single organisation, it may be appropriate to deploy it in such a way that it can only be accessed via the organisation’s internal network (Simsek and Veiga 2001). Controlling and limiting the availability of a web survey also serve to prevent abuse. Though the likelihood of this is minimal for most surveys, this is of particular importance if the survey concerns topics considered to be controversial, sensitive or high-profile. The potential for abuse has been illustrated several times by the abuse of online polls – a different but definitely similar scenario. Time magazine’s 2009 “Time 100” poll was manipulated with enough sophistication to arrange the top 21 results (Schonfeld 2009). A similar hoax in 2010 targeted singer Justin Bieber; an online poll being abused in order to add North Korea as a destination in his first world tour (Emery 2010).

It is worthwhile at this point to acknowledge that numerous applications and Web-based services exist, which can be used to create and administer web surveys. Modern products offer high degrees of customisability and sophistication, allowing a web survey to be created and administered with minimal technical knowledge. While reviews or recommendations of these are not included in this paper (interested readers see for example, Thompson, Surface et al. 2003; Wright 2005 and http://websm.org/), the issues discussed here can serve as criteria by which to assess such applications or services. In addition, one must consider factors such as the cost and the confidentiality, security and ownership of data when determining their suitability (Lyons, Cude et al. 2005; Wright 2005; Greenlaw and Brown-Welty 2009). Despite the availability of these products, various factors often make a purpose-built web survey a necessity.

A final consideration pertinent to the overall design of web surveys is that of the processing and storage of responses. The ability to store response data directly in a database or spreadsheet format is a definite advantage of administering surveys electronically, and this can be maximised with a well-designed web survey (Lyons, Cude et al. 2005; Gonzalez-Bañales and Adam 2007; Greenlaw and Brown-Welty 2009). A web survey is essentially a form, and as such must be processed like any other Web-based form once submitted. Although it is possible to create a form which simply eMails the response to the researcher, entering the response into a database is a much more effective method. This can be facilitated by ensuring that the names of the form elements such as text fields and radio button groups correspond to the database column names, allowing simple and generic form processing scripts to be created. Care should be taken to ensure that web surveys are submitted and stored in a secure manner, utilising appropriate encryption (Lyons, Cude et al. 2005). Survey responses in a database can be queried using SQL (Structured Query Language), offering a means of rapidly extracting and visualising meaningful information from the data. While SQL is unlikely to be the sole tool used for analysis, it is an eminently readable and writable language that can easily be used to explore a data set. The purpose and potential value of the sample SQL statements in Figure 1 should be largely self-evident, even to those with no knowledge of the language.

| SELECT AVG(likert_1), AVG(likert_2) FROM responses WHERE study_mode = 'OC' AND age_bracket >= 2 GROUP BY age_bracket |
| SELECT age_bracket, AVG(likert_3) FROM responses WHERE study_level = 'PG' |

Figure 1: Sample SQL statements used to explore survey responses in a database

To allow effective querying, the values submitted in a web survey should be numeric wherever appropriate. For example, a five-point Likert scale question ranging from Strongly Disagree to Strongly Agree and including a Neutral response should be stored as -2 to 2 with 0 representing Neutral, rather than storing responses textually. This applies to all ordinal values. A database can
easily be exported into a spreadsheet or other formats supported by statistical analysis software as needed.

This section has mainly discussed issues that aim to improve the quality, correctness, usability and reliability of data gathered via web surveys. The recommendations are overarching – benefiting the researcher directly, rather than via the improvement of the respondent’s experience. The following section focuses upon issues that improve the respondent’s experience.

3. Structure, presentation and question layout issues

This section discusses issues of structure, presentation and question layout within web surveys that can influence the respondent’s experience. Regardless of the medium, a survey that is clear, convenient and usable is of more appeal to potential respondents, resulting in higher response rates and potentially better quality data. The author feels it is important to make all relevant information available with the survey itself. While eMail or other methods may be used to introduce the research and request participation, all pertinent information should also be available with the web survey. While this may result in some redundancy, it ensures that respondents do not need to search outside of the survey Web page for any instructions or information they may need. While minor, inconveniences such as these have the potential to deter a respondent – particularly if they encounter the inconvenience before beginning the survey, when they have not yet invested any time or effort into its completion.

The issue of survey length has been addressed in numerous pieces of academic literature, often finding that longer surveys are not only less likely to be completed, but also that “questions asked later in the questionnaire bear the risk of producing lower quality data, especially if they are in open format or in long grids” (Galesic and Bosnjak 2009). Based on such findings (Herzog and Bachman 1981; Gonzalez-Bañales and Adam 2007), two recommendations that are applicable to surveys in any medium are worth repeating. Lengthy surveys should be avoided if possible, and the ordering of questions should be considered in longer surveys – acknowledging that the quality of responses to questions towards the end may be lower than that of those near the beginning (Vicente and Reis 2010). Hence, the most important questions should be placed towards the beginning of a survey. A further consideration regarding the ordering of questions arises from Shropshire, Hawdon and Witte (2009), which found respondent interest to be a significant factor in early termination of a questionnaire. Hence, questions likely to be of highest interest to respondents should appear early in the survey. These two factors should be complimentary in a well-designed survey; important questions should be of interest to a respondent. Two further recommendations can be made for lengthy web surveys in particular. Firstly, the design and presentation of the survey can be tailored to make a long survey more manageable for respondents. One method is to divide the survey into sections and display one section at a time (Figure 2).

This serves to avoid overwhelming the respondent with a large number of questions on one page and the need to scroll, which have been found to influence item non-response and survey non-completion rates (Couper, Traugott et al. 2001; Schonlau, Fricker et al. 2002; Couper 2008; Toepoel, Das et al. 2008; Vicente and Reis 2010). Research (see for example, Tourangeau, Couper et al. 2004; Dillman, Smyth et al. 2009; Toepoel, Das et al. 2009) has found that respondents see questions on the same page as being more closely related, having a slight impact upon the intercorrelation of responses. This should be taken into account when determine if and how to divide a survey into multiple pages. An overall progress bar and section-by-section validation are advisable if a survey is divided into sections. Yan, Conrad, Tourangeau and Couper (2010) examine the use of a progress indicator in web surveys, concluding that it encourages completion if it accurately reflects the expected duration of the survey, particularly if the survey is short. This further emphasises the need to minimise the length of web surveys and to provide information such as its expected duration. Other research regarding the use of progress indicators in surveys divided into sections is generally supportive, however findings vary and the effects of progress indicators are often minor (Couper, Traugott et al. 2001; Crawford, Couper et al. 2001; Vicente and Reis 2010).
The second recommendation specific to lengthy web surveys is giving respondents the ability to save their progress and resume the survey at a later time. This is recommended in Kaczmirek (2005), who states “Do not introduce problems in your online questionnaire which would not occur in a paper and pencil questionnaire.” While implementing the ability to save and resume survey progress is likely to be beyond the technical capabilities of a person otherwise capable of creating a web survey, the feature is common amongst applications and Web-based services that allow surveys to be created.

Unlike mail surveys, web surveys can adapt to a respondent’s answers to questions on the fly. In a mail survey, conditional questions are typically implemented via instructions to the respondent – for example, “If you answered ‘No’ to this question, continue to Question 12”. Using Web-based scripting languages such as JavaScript, web surveys can implement conditional questions in a more effective manner, showing or hiding parts of the survey based on the answers to questions (Oppenheim 1992; Fricker and Schonlau 2002; Gonzalez-Bañales and Adam 2007). In addition to helping to ensure the correctness of data, this serves to “reduce the length of a survey to the individual minimum” (Kaczmirek 2005). Though JavaScript can improve the implementation of web surveys, it is inadvisable to rely upon it heavily. Ideally it should be possible and convenient to complete the survey without requiring the respondent’s browser to have JavaScript enabled. If the survey cannot be completed without JavaScript, potential respondents with JavaScript disabled should be informed of its need when they attempt to access it.

The following recommendations concern the presentation, layout, accessibility and formatting of web survey questions. While some of the recommendations are relatively minor, they are all worthy of consideration when creating a web survey or deciding which survey-generation product to use. Research such as that of Christian, Dillman and Smyth (2007) underpin the potential effects of seemingly minor issues in web survey presentation, finding that the relative size of text fields resulted in an eight percent increase in correctly formatted responses and the use of symbols rather than words as labels resulted in a seven percent increase. In order to avoid excess technical detail, the recommendations are presented in dot-point form. The term “form elements” refers to text fields, radio buttons, checkboxes and other form components used to allow people to respond to questions in web surveys. A number of the recommendations concern the accessibility of web surveys. The need to make the Web accessibly to those with impairments or disabilities is an increasingly significant issue, and one which should be taken into account by researchers – particularly if the research concerns a demographic with a higher proportion of impairments, such as the elderly.

- Make use of space to clearly separate questions from one another. If utilising a grid-based layout, often used to present series of Likert scale questions, use alternating background colours to clearly align question text with the appropriate form elements.

Figure 2: Example of a long survey divided into sections

The following questions concern your current enrolment.

8. What course or degree are you currently enrolled in?:
   Bachelor of Computer Science

9. What is your current level of study?:
   Ⓟ Undergraduate ⓦ Postgraduate

10. Are you currently a full-time or part-time student?:
    Ⓟ Full-time ⓦ Part-time

11. What is your current mode of study?:
    Ⓟ On Campus ⓦ Online / External ⓦ Mixed

12. Which mode of study do you typically prefer?:
    Ⓟ On Campus ⓦ Online / External ⓦ Mixed

Click “Next >>” to continue.

<< Previous Welcome Part 1 Part 2 Part 3 Part 4 Part 5 Part 6 Finish Next >>
Use spacing to ensure that radio buttons and checkboxes are clearly associated with their labels, preventing ambiguity and confusion. Furthermore, use the “label” tag to increase the clickable region of form elements. These points are illustrated in Figure 3.

Figure 3: Likert scale with ambiguous spacing and no labels (A) and Likert scale with clear spacing and labels (B). Clickable regions have been highlighted.

- If there is insufficient space to repeat labels next to each radio button or checkbox, consider using the “title” attribute in the “input” or “label” tag. The content of this attribute will appear as a tooltip when the mouse is over the radio button, allowing respondents to select a response without needing to refer back to other labels.

- Use of the “label” tag also bolsters the accessibility of a survey (Theofanos and Redish 2003; Thatcher, Burks et al. 2006). Forms are one of the largest hurdles faced on the Web by those with impairments or disabilities. A screen reader is the most common piece of assistive technology used by the visually impaired, reading the content of a Web page. When the user focuses upon a form element such as a text field or radio button, the screen reader will read out the label associated with the field. Should a label not exist, the “name” or “id” attribute of the form element is read. This is likely to be generic, such as “question_5”, or abbreviated, such as “oa_sat_lquart” (“overall satisfaction in last quarter”). The former is of little to no help to the user, and the latter will not be comprehensible to the screen reader or user.

- Acronyms and abbreviations are common in surveys, however not all respondents may be familiar with or remember them and screen readers can struggle with their pronunciation. The “acronym” and “abbr” tags can be used to alleviate this, and the “title” attribute can be used to provide further information where needed (Theofanos and Redish 2003; Thatcher, Burks et al. 2006).

- Avoid relying solely upon colours, images or subtle visual cues (such as an asterisk to indicate a required field, “greying out” a disabled form element or using the size of a form element to indicate the desired input length) to convey important information about the survey. Such cues are easily missed or indeterminable by those with low vision, colour blindness, or using a screen reader (Thatcher, Burks et al. 2006).

- The layout of Likert scale responses is explored in Tourangeau, Couper and Conrad (2004), who find that respondents make inferences about the meaning of survey items based on visual cues such as the perceived midpoint of responses, their spacing and order. For example, when nonsubstantive answers such as “don’t know” or “no opinion” were not clearly separated from the Likert scale itself, responses were influenced by the different visual midpoint that this created. These findings should be taken into consideration when designing Likert scales in web surveys.

- If a question requires participants to select from a number of options, use radio buttons rather than drop down lists as radio buttons present the options in a more convenient and accessible manner. Healey (2007) found that drop-down lists result in higher item non-response rates, longer response times and increased accidental changing of responses. Kaczmarek (2005) recommends only using drop-down lists “if respondents know the answer without having to look at all entries”, and to make the default selection is a placeholder such as “select here” so that non-response can be identified.

- Users can navigate form elements via the tab key, with each press bringing the focus to the next element from top-left to bottom-right. The layout of form elements in a survey may cause the default tab order to traverse form elements in an incorrect or unintuitive manner. If this is the case, the ordering can be changed via the “tabindex” in form element tags. The “accesskey” attribute can also be used to allow keyboard shortcuts to form elements if deemed necessary for accessibility reasons. These facilities can also serve to make a form more accessible to those using assistive software such as a screen reader, allowing the form to be navigated effectively without the use of a mouse.

- On the spot validation of responses is possible in web surveys, via scripting languages such as JavaScript. While it is possible to enforce strict validation criteria such as only allowing digits to appear in a field for a phone number, it is not always advisable to do so – for example, the
respondent may want or need to include spaces, parentheses or a plus sign in a phone number. When faced with overly strict, inappropriate or frustrating validation requirements, respondents may invent data, attempt to circumvent the validation, or discontinue the survey entirely (Best and Krueger 2004; Kaczmarik 2005; Christian, Dillman et al. 2007; Gonzalez-Bañales and Adam 2007; Vicente and Reis 2010). Consider including an “other” option in closed-ended questions, and allow respondents to provide open-ended comments (List and Goslin 2002). It is also worthwhile reiterating that reliance upon JavaScript is not recommended.

Failing to take heed of issues such as these can lead to web surveys which are unclear, confusing and frustrating, often resulting in discontinuation or by respondent, or lower quality data. By taking these recommendations into consideration, researchers can design, develop and implement an effective web survey, or select a product which will allow them to create one.

4. Conclusion

There is much academic literature which addresses the design and implementation of effective surveys. Of that, a relatively small proportion discusses issues specific to the increasingly prominent medium of Web-based surveys. Web surveys present unique opportunities and challenges not applicable to mail-based surveys, and care must be taken to ensure that web surveys are developed and implemented in an effective manner. This paper has discussed numerous issues, primarily technical, that should be taken into consideration in regards to web surveys. A number of recommendations have been made concerning the accessibility of web surveys, to raise awareness of the importance of ensuring that a web survey can be completed by someone with an impairment or disability.

Issues pertaining to the overall design, delivery and administration of web surveys were covered first. Email-based surveys are advised against, and the importance of appropriate advertising, unique identifiers and data format were discussed. The recommendations in this section seek to give researchers overarching guidance towards creating web surveys that are more likely to produce data of high quality, correctness, usability and reliability. Next, a number of recommendations were made regarding the structure, underlying code, presentation and layout of questions in web surveys. These focus upon increasing the clarity, convenience, usability and accessibility of web surveys, promoting higher response rates and higher quality data. Many potential respondents to a web survey are not overly motivated or determined to complete it, and hence it is imperative to structure and present a web survey in a way that encourages convenient completion and invests the respondent early to discourage discontinuation of the survey.

As the world’s population becomes increasingly made up of “digital natives”, the Internet has emerged as a key medium for self-administered surveys. The sophistication and availability of modern applications and Web-based services for the creation of web surveys has also made the medium more accessible to researchers in all domains, furthering their use. The recommendations made in this paper assist researchers both in manually developing a web survey and evaluating survey creation products. A well-developed web survey provides both the researcher and respondents with a clear, accessible and highly usable mechanism, minimising the confusing and frustrating elements that may lower response rates or the quality of data.

References


Generating a new Interview Method by using Sensing Technology to Assess Human Emotions

Yayoi Hirose¹, Kiyoshi Itao² and Tomohiro Umeda³
¹Japan Science and Technology Agency, Tokyo, Japan
²Tokyo University of Science, Tokyo, Japan
³Nara Women’s University, Nara, Japan
hirose@sen.jst.go.jp
itao@rs.kagu.tus.ac.jp
umeda@cc.nara-wu.ac.jp

Abstract: This study aims at generating new interview methods for obtaining more detailed information regarding human emotional factors by utilizing sensing technology. It can improve the weaknesses of qualitative research past discussions have pointed out, and develop the validity of collected data and more objective analysis of collected data in qualitative research. As the first step for a new research method, the study uses two types of sensing device which assess the emotional condition of an interviewee. The first device is ST technology, voice analysis software and a system of emotion estimation. This device using voice analysis defines what emotional condition the interviewees have. When the interviewee makes a statement, the ST technology can investigate his-her emotional condition, such as whether the interviewee stated it disappointedly, delightly, or angrily, while the conventional coding simply relies on text data. The second device is WHS-1, portable sensing device. This device investigates whether the interviewee stated certain things in a relaxed or stressed condition by measuring heart rate and analysing the condition of autonomic nerves. The study finally adopts both devices to precisely assess the interviewee’s emotional condition, and demonstrates that the two devices enable the researcher to obtain closer view of the interviewees. The study suggests that generally the two types of sensing device can play a supportive role in analysing emotional factors for interview research. While the researcher can only ascertain stressful or delightful factors based on coding analysis, using sensing devices enables the researcher to identify how stressful or joyful the interviewee is, and why they are in this state, ie. are they stressed due to anger or sorrow? It is expected to enable the researcher to more deeply consider the reason the interviewee is in a certain emotional condition, and also lead to contextual or theoretical discussion.

Keywords: interview method, qualitative research, sensing technology, emotional factor

1. Introduction

This study aims at generating a new interview method of qualitative research, by assessing interviewee’s emotional condition in more detail than the conventional coding method. This method is expected to be used for improving various kinds of communities, and in particular, can be expected to help improve social interactions between elderly people and care providers. Several countries are now faced with an aging problem. In particular, the Japanese population took a downward turn in 2005, and Japanese society is now comprised of 21 per cent of over 65 year-old people, which Japan has never experienced. The working population of Japanese society is expected to decrease rapidly, and society will be largely occupied by elderly people in the near future. Discussion is now being held on how to generate a sustainable super-aging society. Due to this situation, the national government initiated public nursing care insurance ten years ago, and a large number of senior businesses have been generated, including elderly care services, hospices, nursing home services, and home help services. However, the care business for the elderly has the severe problem that a lot of care workers quit their jobs because of psychological stress. One of the reasons is that care workers are not motivated to provide better care services because they cannot interact with the care receivers very well. For example, care workers cannot understand whether the elderly really enjoy their current circumstances. The elderly who can express their intentions can enjoy their care services, such as listening to music. However, the care receivers may become unwilling or hesitant to make requests of those providing care, as they do not want to be thought of as troublesome or a nuisance. For example, the care receivers sometimes may not feel comfortable saying, “I want to go to bathroom. Please help me.” In order to investigate this factor, interview or survey is regarded as an important method. However, several senior people cannot necessarily express their own opinions, and especially senior people who require nursing care might have difficulty in speaking.
The care business for the elderly suffers the severe problem that a lot of care workers quit their jobs because of psychological stress. It is a huge problem for Japan, which expects to become a super-aging society in the near future. As the aging population grows rapidly, the short-handedness of care workers will also grow into a serious problem. Japanese society now needs to deal with this problem. The survey says that one of the main reasons that care providers feel stressed is that they cannot understand whether they are providing appropriate care services to care receivers and whether care receivers are really delighted by their services (Care Work Foundation, 2005). While Veterans take a seat-of-the-pants approach, a lot of care providers of several years experience still wonder what to do. On the other hand, as a Japanese sociologists points out based on her research (Ueno, 2011), care receivers may also hesitate to make requests of care workers. In other words, the care service needs to solve both sides of the misunderstandings and discrepancies, and promote human integration. Based on this situation, this study, considering the needs for improving the interaction between elderly people and their care providers, discusses the possibility of new methods of qualitative research by analysing emotional conditions of speakers. First, the study reviews past discussions on the analysis of emotional conditions of interviewees and digital tools for interviews and analysis. Next, the study introduces two sensing devices, ST and WHS-1, and demonstrates how these express emotional conditions when conducting interviews. Finally, the study conducts mixed analysis, conventional coding method, ST, and WHS-1, and shows how the two sensing devices enable the researcher to more accurately analyse interview data. In concluding section, the study summarizes the benefit of using the sensing devices and outlines necessary further development.

2. Research method

2.1 Analyzing emotional factors

In providing elderly care services and developing interaction between the elderly and care providers, understanding human emotional factors is paramount. Past literature claims that the qualitative approach is appropriate for an inquirer involved in sustained and intensive experience with participants, as it focuses on participants’ perceptions and experiences by using multiple methods that are interactive and humanistic (Creswell, 2003; Fraenkel & Wallen, 1990; Merriam, 1998; Thomas, et al., 2005). In the process of qualitative research, several new aspects emerge by interpreting and analyzing data. This is because qualitative research develops a description of an individual or setting, and interprets and concludes personal or theoretical new meaning (Bogdan & Biklen, 2003; Thomas, et al., 2005). Thus, qualitative research requires the interpretation of the settings from multiple perspectives. Based on these arguments, in order for care providers to consider and sometimes create new services favorable for various types of elderly person, the qualitative approach is appropriate as it clarifies favorable care services for each elderly person.

In order to conduct qualitative research, the researchers collect data based on open-ended observations, interviews, documents, which include e-mails, scrap books, and audiovisual materials, including photographs, videotapes, computer software and film (Creswell, 2003). Among these data, qualitative researchers have often used written documents of interview data when they analyze when and how people think (Akiyama, et. al., 2006). When analyzing the interview data, qualitative researchers analyze the documents line by line, following the approach outlined by Strauss and Corbin (1998) (Easterby-Smith, 2008). Even when the voice of interviewees is recorded and researchers analyze the emotional state of the interviewees, analysis is mainly based on written records (Remenyi, et. al., 1998).

Past discussions point out there is a bias the qualitative researchers bring to the study (Creswell, 2003; Remenyi, et.al., 1998). The analysis of written accounts is not enough to understand interviewee’s subjective experience, based on the argument that qualitative researchers seek to understand the context or setting of the participants through “visiting this context and gathering information personally” (Remenyi, et.al., 1998). For elderly care, researchers need to consider people’s contextual differences, as some of the elderly may hesitate to clearly request what they really want to do. Sometimes, even if they do not enjoy care services, they may say, “I am enjoying this service.” The problem is when a care receiver makes this statement, the interviewer cannot make the assessment that they are not as happy as they maintain. Therefore, researchers need to seek more multidisciplinary research methods in order to objectively analyze what people really think based on a multilateral approach. This problem may affect the subsequent interpretive processes where the researchers aim at understanding why certain things occur. In order to validate the accuracy, qualitative researchers have used member-checking and triangulation of different data sources. However, even these methods cannot
help avoiding subjective biases completely. When a researcher considers emotional factors in detail, more objective methods based on scientific research can provide further validation of accuracy.

There is a research tool for measuring human emotions, such as AdSAM. However, this tool is for marketing research, and mainly for measuring a large range of human emotions (AdSAM, 2010). In order to analyze human emotional factors, researchers mainly use written documents. For example, AsSAM can be used when analyzing whether targeted consumers make purchasing decisions and how they feel when they see an advertisement. On the other hand, nursing care for the elderly requires one-to-one based care and offers the elderly emotional support.

Zarit (1980, 1990) creates a specific scale for a care provider’s burden, “Zarit Caregiver Burden Interview: ZBI”, which includes physical, psychological, and economical burdens of care providers brought by supplying care services. This scale is composed of 22 questions regarding the burden of care services, and questionees choose one of the five scales. It is useful to clarify what kind of burden a care provider has and what factor is the most severe for care providers. However, in the super-aging society, interaction between care providers and care receivers is paramount (Ueno, 2011), and it is necessary to understand when and how people feel at a specific situation, such as during maintaining hygiene and meal time. This will bring about a more sustainable aging society, where both understand each other's feelings and are motivated to communicate.

Thus, in order to provide sustainable care services, it is necessary to collect data on how people feel in various situations. One-to-one care based on specific situations requires consideration of what senior people feel is different depending on individual experiences and emotional conditions which are socially and historically constructed through experience. For example, it depends on the social value they belong to whether senior people can easily express what they want. In some societies which value low-assertiveness, such as Japan, senior people may not be encouraged to express their needs for care services (Hartog, 2004). They may feel stressed to listen to music which they are not interested in every day and is different from their conventional life style.

2.2 Digital tool

Qualitative researchers have been discussing the utilization of various types of technological devices. Brown (2002) introduces multimedia databases which can capture, store and retrieve digital data. Several articles point out Maxqda, ATLAS-ti NUD*IST and NVivo as tools for digital data analysis (Bryman & Bell, 2007; Easterby-Smith, 2008; Lewins, et al., 2007). Knoblauch et al. (2008) discuss how the development of information technology enables researchers to generate new visual research methods, including interpretative analysis of video and photography.

However, there is little discussion among qualitative researchers regarding analysis of human emotional and physical condition by technological devices. Although audiovisual materials are convenient for researchers to directly understand their reality, problems have been pointed out, such as difficulty of interpretation (Creswell, 2003). This study aims at analyzing human emotional factors by adopting two technological devices. Both of them use advanced sensing technology and aim at analyzing emotional condition by collecting data inside the human body, such as human voice and heart beat during the interviews. The devices are developed based on scientific research. Thus, assessment of emotional condition is based on more objective data rather than a researcher’s assessment, which is based on the researchers’ skill.

3. Research method and analysis

Based on these arguments of the preceding section, this study aims at analyzing “how they felt?” more objectively by conducting interviews. The study focuses on the stages of collecting data and interpreting the collected data, and is just the first stage of generating new interview and analytical method. The study may enable researchers to offset the weakness of conventional qualitative researchers, which are pointed out in past discussions (Merriam, 1998), as it can enable the collection of more valid interview data and analyze interview data in a more objective way. In this section, researchers conduct case studies by using ST and WHS-1 and compare the interviewee’s statements with their emotional conditions and support the information which written documents cannot assess. The study selects the interview data of general people who can express what they feel, as statement from interviewees who cannot express their opinions cannot be compared to their emotional conditions.
3.1 Sensibility technology ST emotion (ST)

The first device is a product of AGI Japan Inc. “Sensibility Technology ST Emotion”, voice analysis software and system of emotion estimation, which measures emotional parameters in speech. This software assesses what emotional factor is prominent, anger, joy, sorrow, and calmness when a person states a certain thing. The software uses power, fundamental frequency (F0), and their transition as well as speed and intonation of speech to analyze the voice, as F0 and power refer to the vibration of vocal codes and may be an expression of natural human emotion (Mitsuyoshi et al., 2007). The system detects changes of feeling or emotion that a person does not notice, and it can detect more than 70% of human emotion (Tokuno et al., 2011). The study analyzes interview data using ST technology and compares it with the analysis based on written documents of interviewee’s statements. The researchers obtained guidelines from AGI for interpreting the outcome of the ST analysis. Based on Mitsuyoshi (2008) and the guidelines, the researchers interpreted the interview data.

3.1.1 Analysis based on coding

The researchers selected three interviewees. The first person is a male care worker, with a wife and 8 month old girl. The second person is his wife, who used to be a care worker. She quit her job after marriage and childbirth. The third person is a female care worker, with a husband and a child. They were interviewed about how they feel about their present life.

On analysis of how a care provider and his wife think about their present situation, one theme emerged for each person,

1. Satisfied with his present life: interviewee 1
   1-1. Satisfied with their daughter’s smile.
      *Her smile is really great. I like both the smile of the elderly and children* (Interviewee 1)
   1-2. Satisfied with the elderly smiles.
      *I like the elderly’s smile. I see their smiles a number of times so I cannot quit my job. I like their smile.* (Interviewee 1)

2. Anxious about the future: interviewee 2
   2-1 concerned about their daughter in the future: Interviewee 2
      *I am anxious about whether I will be able to do financially for my daughter what my parents did for me.* (Interviewee 2)

3. Satisfied with her present life: Interviewee 3
   3-1: Satisfied with her job
      *I like the elderly.* (Interviewee 3)
      *I enjoy my work. I think this is my true vocation, because I enjoy my work.* (Interviewee 3)
      *My boss asked me, “Do you like your job?” and I said, “Yes. I like it.”* (Interviewee 3).

Based on this initial coding, a researcher may think that interviewee 1 is satisfied with his present situation as he can look at his daughter’s and elderly people’s smiles, while his wife is financially concerned about her daughter because of her husband’s low income. Interviewee 3 is satisfied with her present situation as she likes interaction with the elderly.

3.1.2 Analysis based on the ST technology

On the other hand, the ST displays the two people’s feelings in more detail. Although interviewee 2 certainly talked about her concern about her daughter in the future, her emotional condition was “CALM” during this statement. Rather, she clearly presented “SORROW” when she explained about their financial situation.
Compared with people who graduated from university and work at a business office, in our work, care providers must work much harder. (CALM) But our income is too low. (SORROW).

We do not have discretionary money after childbirth. (SORROW). Our common hobby is bringing up our child. (CALM)

The ST can propose different messages from the analysis of interview documents. Based on the analysis of the written interview documents, a researcher cannot understand whether the interviewee 2 just told the truth about her financial situation, or is unsatisfied with her present life. On the other hand, the ST suggests that his wife is more disappointed about the situation due to his husband’s low salary, as she does not have discretionary money. She certainly may have strong feelings about the low income of care providers despite their hard work. Tokuno et. al. (2011) show that strong stress increases the feeling of “sorrow” and decreases the feeling of “joy.” Therefore, based on the ST analysis, it can be said that she may be stressed with her present financial situation rather than her daughter’s future. On the other hand, interviewee 1 did not display “SORROW” during the interview, although he explained about his low salary. This implies that while his wife is highly concerned about her husband’s low salary, he is not so disappointed and takes it coolly.

The ST also has the possibility to clarify the difference in human satisfaction between Interviewee 1 and Interviewee 3. Interviewee 1 consistently presented “CALM” whatever he was asked. Even when he explained about his job and his favorite elderly smile, his emotional condition was calm. This implies that he is very relaxed with the smile of the elderly and his daughter. The main reason that he continues to be a care worker was that he can be relaxed with their smile.

Interviewee 3 shows more “JOY” for the elderly. In particular, she felt delighted when she talked about how she decided on her job. She explained that as a high school student, she wanted to be a children’s nurse, and changed her mind when she experienced voluntary practical training for elderly care. During this talk, she displayed JOY to a certain degree. After that, when she stated “I also like the elderly,” she displayed a stronger joyful condition. This suggests that she may really like the elderly and her job. Furthermore, she displayed “JOY” when she stated, “I realized there are many misunderstandings about my job,” and talked about the process of how she attained her work-life balance from the initial biased idea that it is impossible to balance both. Based on analysis of documents, it is not clear whether she is satisfied with her working style, while the ST displays she is extremely satisfied with her work-life balance. This may imply that the main reason that she continues her job is that she is more motivated by interaction with the elderly under a work-life balance.

Table 1 summarizes the analysis of the three interviewees.

**Table 1: summary of analytical difference between coding and ST**

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Analysis based on statement</th>
<th>Analysis based on the ST device</th>
<th>Final analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Satisfied with his present life. I like the elderly’s smile. I see their smiles numbers of times so I cannot quit my job. I like their smile.</td>
<td>Neutral with his low salary. We have not been purchasing clothes (CALM).</td>
<td>He is satisfied with his present life, and not so stressed with his family’s financial situation.</td>
</tr>
<tr>
<td>2</td>
<td>Concerned about her daughter in the future. I am anxious about whether I will be able to do financially for my daughter what my parents did for me.</td>
<td>Stressed with the husband’s low salary. We do not have discretionary money after childbirth. (SORROW).</td>
<td>She is not satisfied with her present life. She is stressed with her husband’s low salary and concerned about her daughter.</td>
</tr>
<tr>
<td>3</td>
<td>Satisfied with her present life, as she likes interaction with the elderly. I think this is my true vocation, because I enjoy my work.</td>
<td>Feel joyful to do her job. I also like the elderly. (as well as children)(JOY) I realized I was being biased towards my job. (JOY)</td>
<td>She is satisfied with her present life, as she enjoys her job and her present work-life balance.</td>
</tr>
</tbody>
</table>
3.2 WHS-1

The data shows that the ST software sometimes includes some errors. For example, when interviewee 1 looked at his daughter, the ST interpreted his sigh as “ANGER” and then his laughter as “JOY”. Based on his comments and smile, it is obvious that he feels JOY when he looks at his daughter. The ST also sometimes mistakenly interprets “STRESS” for a joyful condition, when people laugh in a loud voice. Thus, the ST cannot analyze all of the voice data. In order to analyze emotional condition during talking more precisely, this study adopts the second device, WHS-1, which is a small-sized portable sensing device. It measures human biological information, such as heartbeat, which includes wave shape and cycle of human heartbeat, temperature of human body surface, and triaxial rate of acceleration, and monitors a human biological condition based on specific analysis by PC software.

This sensing device is produced based on the past research which shows that LF/HF, Low Frequency / High Frequency, conducted by spectral analysis of heart beat variability, is high when sympathetic nerves activate (Takada, et.al., 2005). As this device mainly clarifies a general trend of emotional condition over several hours, while the ST assesses the emotional condition phrase by phrase, it has been mainly utilized for analysis of severe depression. Therefore it is necessary to clarify to what a degree this device shows emotional condition over a short period of time, such as 2-3 minutes. In order to assess the usefulness of this device as a complimentary source of interview data, the researchers measured a person’s heart beat and worked out LF/HF figures during the conversation and analyzed human emotional conditions. The examinee is female, 23 years old. She wore the sensing device, and talked with the researcher. The researcher sometimes had a very pleasant conversation with her, such as her hobbies and sometimes strongly ordered her to do so much work within a very short time.

Table 2 shows the difference of LF/HF figure among her calm condition, relaxed talking and stressful talking. It shows each average figure during a temporal sequence of three minutes. The comparison shows that the LF/HF figure is relatively higher at the stressful condition than relaxed talking and calm condition. During the calm condition, the LF/HF is lowest. The data shows that when the examinee enjoyed relaxed talking, she was in a certain level of excited condition. As the emotional trend is different between the three conditions even within 3 minutes, it can be useful to complement the analysis of interview documents.

Table 2: Comparison of LF/HF figure between calm, relaxed, and stressful conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Calm condition</th>
<th>Relaxed talking</th>
<th>Stressful talking</th>
</tr>
</thead>
<tbody>
<tr>
<td>average</td>
<td>3.36</td>
<td>6.56</td>
<td>8.52</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.23</td>
<td>2.95</td>
<td>3.73</td>
</tr>
</tbody>
</table>

3.3 Mixed analysis

Finally, the researchers conducted an interview regarding when a care provider felt stressed and relaxed and analyzed the interview by coding, the ST device, and WHS-1 device. The interviewee is a female, 43 years old, used to be a care provider. The researcher asked her to look back at her job experience and explain about her stressful and relaxed experiences. Based on the initial coding, several codes which present stressful occasions and relaxed occasions came out. The researchers compare these codes with the ST results and WHS-1 results. Based on these results, the ST and WHS-1 well support stressful factors based on the coding. For each code, the result of the ST analysis will be explained in the following sections based on AGI’s guidelines and Mitsuyoshi (2008). Transition of LF/HF measured by WHS-1 during the interview is shown in Figure 1, while it is impossible to show the transition of all figures of the ST results due to space limitation. Figure 2 shows one part of the ST result. By and large, the ST displays “ANGRY” very frequently. This implies that the interviewee was very stressed about her experience as a care worker. During the interview, she looked back at her experience and presented stress, complaints and anger. On the other hand, the ST rarely displayed CALM while she was talking, while WHS-1 sometimes displayed a relaxed condition of the interviewee (Figure 1). The following explains several stressful and relaxing factors based on three types of analysis.
1. Stressful factors

1-1 Required heavy labor in order to avoid fatal accident despite insufficient physical ability

Giving a care receiver a bath is the heaviest labor among care work, such as standing the elderly and moving them to the wheelchair. I was scared to think if I mistakenly broke their clinical instruments, including the tube of respiratory apparatus. I was really scared to give these care receivers a bath. I was so scared and wondered what to do if I broke the instruments. I cannot do anything to avoid these types of accidents as the elderly are so heavy.

Figure 1: Transition of LF/HF during the interview

1-2 Care receivers’ negative reaction to her services

Giving the care receivers who hate taking a bath was really difficult. Some of them strongly refused and sometimes told lies that they took a bath yesterday. But their bodies smells badly. So I tried to persuade them to take a bath. But if they completely refused, I reported, “they refused.”

Figure 2: ST result which displays a stressful condition

The first stressful factor that came out during the interview was heavy labor required to avoid a fatal accident for the elderly. WHS-1 also displayed high LF/HF during this section of over 20 seconds (zone 1 in Figure 1). The second half of Figure 2 shows the ST result of this section. During this section, based on the ST, she felt angry with high excitement, but her joyful feeling was also sometimes high. Based on the AGI guideline, it can be said that the interviewee felt joyful as the interviewee listened to her and understood her difficulty. Considering the interviewee’s statement it implies that she wanted to emphasize heavy labor, such as giving care receivers a bath.
During this section, the ST displayed ANGRY and sometimes JOY with a certain level of excitement. Considering the interviewee’s statement and AGI guideline, this implies that the speaker may be a little bit fed up with the situation and wanted the interviewer to understand her hardship. The LF/HF figure is also extremely high for 25 consecutive seconds (zone 2 in Figure 1). Thus, based on coding, it can be said that this care worker is stressed with the difficulty of persuading care receivers to do necessary things. Furthermore, the ST and WHS-1 suggest that she may also be fed up with the elderly.

1-3 human relationship with incompatible colleagues in her work site

Umm...human relationship in the care home... Yes. There are several troublesome care providers. I think they started care provider jobs because they cannot be working people. I was fed up with them....One of them talked about herself that she used to resist going to school in a meeting room where there were a lot of complete strangers existed.

During this section, the ST displayed that the speaker felt ANGER and SORROW. This implies that the speaker sometimes felt angry and sometimes got bothered by the situation. The LF/HF figure was also extremely high and this condition continued over one minutes while talking about this topic (zone 3 in Figure 1)

1-4 Need to take a responsibility for the accident that could not be avoided

Even if I just observed the situation where care receiver exercised and fell down, I needed to take responsibility, and was required to write a report about why this happened. I was required to watch the care receiver at that time. But my colleague stood by me and said “Ah it cannot be avoided.”

Based on the analysis of the interview document, the next stressful factor for her can be necessity to take responsibility for an accident which could not be avoided. During this section, the ST displayed that the speaker mainly talked with ANGER, and sometimes also mixed with JOY. Based on the AGI analysis, this may imply that the speaker talked with anger and self-assertiveness. The WHS-1 shows that as LF/HF is very high and the high level continued for over 90 seconds (zone 4 in Figure 1), she may feel stressed.

1-5 not understanding care receiver’s physical condition

I was disappointed when a care receiver discharged his bowels in the bath. I needed to restart the whole process of giving a bath from scratch, and washing the bath was also heavy work. I felt if they gave me even a small sign. Some of the elderly did this as a joke....I do not think they were malicious...ummm.

During this section, based on the ST, the care worker felt angry and subsequently sorrowful. The LF/HF figure was also high and the high level continued for one minute (zone 5 in Figure 1) This implies that the care worker felt angry that the care receiver did what the care worker did not expect, but was stumped about not understanding the care receiver’s condition, which brought about this situation.

2. Relaxing factors

2-1 communication with the care receivers

I was happy when I interacted and communicated with the elderly. The most delightful thing was that I was sometimes able to communicate with the care receivers, even if they suffered Alzheimer’s disease. As they suffered from the disease, they sometimes forgot what they said just after speaking. But sometimes, I heard what I did not know, like a war story.

During this section, LF/HF remained at a low level for one minute (zone 6 in Figure 1). This means that she was relaxed during talking about her delightful things. On the other hand, the ST suggests that she felt angry and joyful. During this section, she imitated the care receiver’s voice or colleague’s voice. Thus, this imitation may sometimes be interpreted as ANGRY.

2-2 human relationship with compatible colleagues in her work site
When our care center held a barbeque party, we saw a lot of beef on our table, even if the elderly would not eat beef. But after the party, the beef was gone…(laugh) Who ate it! Maybe the male care providers (Laugh)

During this section, the ST suggests that she sometimes felt joy and sometimes angry. But it was obvious for the interviewer that she talked about joyful experiences at her work site. Thus, it can be said that her laugh was interpreted as angry. It was confirmed by the low LF/HF figure for one minute (zone 7 in Figure 1).

3. The result of the analysis

The result of the mixed analysis is summarized in Table 2. The results suggest that using sensing devices enables researchers to obtain a closer view of the reality of the interviewees.

Table 2: Summary of the mixed analysis

<table>
<thead>
<tr>
<th>Initial codes</th>
<th>ST analysis</th>
<th>WHS-1 analysis</th>
<th>The result of the analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1 Stressed with heavy labor required</td>
<td>Wanted to complain about the heavy labor to the interviewer</td>
<td>Highly stressful</td>
<td>Stressfully pleaded for understanding about her heavy labour</td>
</tr>
<tr>
<td>1-2 Stressed with negative reaction to her services</td>
<td>Wanted to complain about her difficulty and fed up with care receivers</td>
<td>Highly stressful</td>
<td>Stressfully made a plea to the interviewer to understand her situation</td>
</tr>
<tr>
<td>1-3 Stressed with human relationship with incompatible colleagues</td>
<td>Felt angry and got distracted</td>
<td>Highly stressful</td>
<td>Stressed and angry about being distracted by the human relationship</td>
</tr>
<tr>
<td>1-4 Stressed with need to take a responsibility for care receiver’s accidents</td>
<td>Felt angry and sometimes self-assertive</td>
<td>Highly stressful</td>
<td>Highly stressful and angry about her responsibility system</td>
</tr>
<tr>
<td>1-5 Stressed with not understanding care receivers’ physical condition</td>
<td>Felt angry and helpless about</td>
<td>Highly stressful</td>
<td>Highly stressed and felt helpless about her situation</td>
</tr>
<tr>
<td>2-1 Communication with care receivers</td>
<td>Includes some errors</td>
<td>Highly relaxed</td>
<td>Felt relaxed on communication with the care receivers</td>
</tr>
<tr>
<td>2-2 Human relationship with compatible colleagues</td>
<td>Includes some errors</td>
<td>Highly relaxed</td>
<td>Felt relaxed with communication with compatible colleagues</td>
</tr>
</tbody>
</table>

When the interviewee made a negative statement about her work in 1-1 code, the ST indicated that the interviewee wanted to complain about her job. This implies that as the interviewee is open-minded to the interviewer, the interviewee expected the interviewer to understand her real situation. Also, considering that the WHS-1 showed that the interviewee was highly stressed, she was never in a calm state and stressfully pleaded for understanding of the fear of causing fatal accidents. In 1-2 code, it can similarly be said that she stressfully made a plea to the interviewer that she was fed up with the difficulty in persuading the elderly to receive necessary services. To understand that the interviewee is open-minded to the interviewer is highly helpful for effective interview research. As Easterby-Smith, et al. (2008) discussed, it is necessary for an interviewer to understand what the interviewee wants to say, and what they do not want to say. Obtaining trust and constructing an open relationship with the interviewee are important but difficult interviewing skills in order to obtain from interviewees the data that an interviewer really wants to know (Easterby-Smith, et al., 2008). In this study, considering the result of the ST device, it can be said that the interviewer was successful in obtaining detailed information on the interviewee.

This analysis also enables a researcher to consider the background context of why the interviewee stated certain things, as the data obtained by the sensor technology is expected to enable the researcher to analyse further why the interviewee stated certain things in certain emotional conditions. For example, the results of ST and WHS-1 suggest that she stressfully felt angry and distracted about human relationships and also self-assertively felt angry about the responsibility system in her organization.
based on 1-3 and 1-4 codes. These imply that she were not satisfied with the management system of her organization. It can also be possible to think that the management of her organization was under stressful conditions. It can be thought that her organization forces the service providers to take all responsibilities and many care providers, may feel stressed. Furthermore, 1-5 code suggests that she was stressed and felt helpless about not understanding the care receiver’s condition. These results may enable the researcher to consider the necessity of improving the management system of care provider’s organization by supporting the care providers to help focus on constructing intimate relationships with care receivers rather than being distracted by human relationships with their colleagues.

On the other hand, the researcher needs to consider the possibility of errors which the devices may cause. For example, regarding delightful occasion, the ST sometimes sent confusing messages to the researcher, such as misinterpretation of imitation of other people’s voice and laughing as ANGRY, while the rest of the statement showed JOY in code 2-1 and 2-2. The WHS-1 supported these, and it enabled the researcher to interpret that the interviewee felt relaxed and joyful about the communication with the elderly and her compatible colleagues.

4. Conclusion

This study discussed the possibility of analyzing emotional factors of interviewees by using sensing devices. In particular, the researchers focused on the stages of collecting data and analyzing interview data and discussed the possibility of generating new methods of collecting data based on higher levels of validity and interpreting human emotional conditions more objectively. The study suggests that utilizing the sensing devices enables more objective analysis regarding interviewee’s emotional conditions, while conventional methods cannot avoid a certain degree of researcher’s subjective bias. For example, when the interviewee complains about their jobs, the analysis results are different between the stressful and calm conditions. When the interviewee angrily complained about their jobs in a stressful condition, the data suggests that the interviewee has a certain ideal situation and finds the present work condition as a problem. On the other hand, when the interviewee complained about their jobs in a calm manner, it can be thought that the interviewee has an emotional detachment and views this as the nature of the job. Using these devices can contribute to the development of qualitative research in terms of strengthening validity of collected data and more objective data analysis. However, considering certain inconsistencies and sample size, further related discussions and detailed data are necessary for future research.

Based on the fact that the sensing devices can assess the human emotional condition in detail, they can be helpful for a wide range of business research including the fields of marketing and customer-satisfaction. Especially in Japan, the devices are expected to be used to investigate people who cannot clearly express their feelings and opinions, such as senior citizens, considering that Japan is approaching a super-aging society. By obtaining the emotional condition data of senior citizens, it becomes possible to generate business which senior citizens really want. Exploring the emotions of those in care in detail will connect the emotional gap between care receivers and care workers, and promote communications between each group. By focusing the service on what senior citizens want, sustainable senior businesses can be generated. High-end technology has struggled to accurately measure human emotional conditions. Further technology development can also be expected.

Acknowledgements

We would like to thank Dr. Mitsuyoshi, the president of AGI, whose comments were of great value for our study.

References

http://www.adsam.com/about/adsams-survey-on-the-american-marketing-associationss-2010-marketing-research-conference (accessed on 17/01/2012)
Care Work Foundation (2005) A Survey Report Regarding Care Worker’s Stress.
Mitsuyoshi, S. Lecture material regarding the ST at the University of Tokyo University (Japanese) (http://www.agi-web.co.jp/docs/Univ-Tokyo.pdf)
The Educational Scorecard: The Start of our Journey

Vivienne Griggs¹, Michelle Blackburn² and Joanna Smith¹
¹Leeds Business School, Leeds Metropolitan University, Leeds, UK
²Sheffield Hallam University, Sheffield, UK

v.griggs@leedsmet.ac.uk
m.blackburn@shu.ac.uk
j.a.smith@leedsmet.ac.uk

Abstract: Against a backdrop of public sector cuts, increasing university fees and high youth unemployment, we are facing challenges in Higher Education to demonstrate the value of our courses. Assessing the value of learning, however, is not straightforward. This paper reports on a study of evaluation processes on a post-graduate, professionally accredited diploma delivered at a selection of post 1992 universities. The driver for the study was a concern that current evaluation processes do not fully demonstrate the value of the course nor take into account the needs of multiple stakeholders. The project included benchmarking University evaluation processes and conducting a dialogue with stakeholders. The study adopted a qualitative management research approach, involving: a review of current practice, comparison with a sample of equivalent courses, and consultation (in the form of focus groups and semi-structured interviews) with a sample of students and employers. The paper discusses findings and proposes recommendations for future evaluation procedures in the design of an ‘educational scorecard’ for the course. Kaplan and Norton’s ‘balanced scorecard’ (1996) concept was adapted to provide a mechanism to represent and balance the needs of different stakeholder groups in the education process. We argue that we cannot truly assess accountability and comparability without engaging a range of stakeholders, not only in soliciting their views on the outcome of the learning but also in the design and implementation of evaluation processes. The ‘educational scorecard’ presented in this paper is developed specifically for the post-graduate diploma in Human Resource Management (PDHRM) at Leeds Metropolitan University and the report concludes with some initial reflections on the benefits of adopting the scorecard methodology. The proposed model is flexible and may be adapted for other HE institutions and courses.

Keywords: evaluation, HRD, higher education, stakeholders, balanced scorecard

1. Introduction

The plethora of research available suggests extensive critique of evaluation of learning in academic settings. Loveland (cited in Haan, 2010) claimed there have been over 2,000 studies in student evaluations of teaching during the past seventy years. Therefore it is inappropriate to suggest a failure of activity. The challenge remains, however, about what, how and when we seek to evaluate. The Course Leadership team for the Post-Graduate Diploma in Human Resource Management (PDHRM) at Leeds Metropolitan University had some concerns about whether the existing evaluation procedures were adding value and fit for purpose. As teachers of HR we examine levels of evaluation (Kirkpatrick, 1998), purposes of evaluation (Easterby-Smith, 1994), return on investment (Phillips and Phillips, 2001) and return on expectation (Anderson (2007)). We assert the importance of and inherent challenges in the evaluation to the learning and development effort. As advocates of this approach however we often fail to demonstrate appropriate practice, at the end of the module we hand out the ‘happy sheet’ (a module evaluation form) focused on teaching rather than learning and assessing satisfaction rather than achievement. The aim of the project reported in this paper was to investigate current evaluation procedures on the PDHRM and consider to what extent they meet the requirements of different stakeholder groups.

It is clear that today’s universities have a wide range of stakeholders all with complex needs. In attempts to create a framework for measuring delivery against stakeholder requirements, researchers (e.g. Karathanos & Karathanos (2005); Beard (2009)), have sought to draw upon the work of Kaplan and Norton (1996) on the ‘balanced scorecard’ and attempted to create a framework that will measure performance of an academic institution. All of these writers have tended to accept the Kaplan and Norton (2006) 4 box framework. Barnth et al (2011) however, sought to create a sector appropriate model albeit driven by a financial management perspective, and specified ‘research’, ‘teaching’ and ‘service’ as the three perspectives for measurement. This paper seeks to establish whether there is an appropriate ‘educational scorecard’ for the post-graduate HRM course through an evaluation of the needs of varying stakeholder groups.

A key challenge for evaluation of a part-time post-graduate, professionally accredited, course is the scope and complexity of stakeholders. Sarrico et al (2010) found that the existence of multiple stakeholders leads to conflicts in evaluation, as different stakeholders may have different perspectives and priorities. The challenge for institutions is to develop an evaluation framework that can accommodate the needs of all stakeholders, while ensuring that the evaluation process is comprehensive and effective. This paper seeks to establish whether there is an appropriate ‘educational scorecard’ for the post-graduate HRM course through an evaluation of the needs of varying stakeholder groups.
stakeholders was often ignored in HE evaluation. It could be argued that we cannot truly assess accountability and comparability without engaging a range of stakeholders in the design and participation of evaluation procedures. Many evaluation procedures focus solely on student feedback but this means that the data collected are not triangulated with that of other sources. Sproule (cited in Nygaard and Belluigi, 2011) goes so far as to argue that the exclusive use of student feedback to evaluate teaching and/or courses is an ‘anathema to the academic mission’.

The current climate for Higher Education presents a challenging time for Universities. The economic conditions coupled with Government policies to reduce public-sector funding but seemingly increase regulation, present both a squeeze on student numbers and a more competitive market place. Chris Woodhead, former chief inspector of schools, claims vice chancellors have ‘turned a blind eye to these issues of quality’ and suggests higher tuition fees will encourage students to demand a better service (Answer the Question, 2012) This exerts pressure on universities and ultimately course leaders and lecturers, who come under greater scrutiny to demonstrate added value at an acceptable cost, and to respond in a constructive and timely manner to student demands. The Chartered Institute of Personnel and Development (CIPD) suggest it is particularly critical in the current economic climate to ensure a good return on investment for learning initiatives as many companies will have increased focus on costs (CIPD, 2011). The majority of students on the course under investigation are part-time, professional students, and therefore we face similar pressure to demonstrate the value of investment in the course to both students and employers.

This paper unfolds as follows: first we cover a brief review of issues concerning evaluation procedures in HE, then the research approach for the study is explained, the findings are then explored for each respondent group under three key themes, a proposal for an ‘educational scorecard’ is presented, and finally we conclude with some initial suggested benefits of the scorecard approach and some suggestions for future research.

2. Evaluation procedures in HE

The CIPD annual Learning and Talent Development survey (2011) found post-course evaluations or ‘happy sheets’ to be the most commonly used method of learning evaluation (93% of relevant organisations) this is also the approach found in academic institutions. This approach in education is a standardized form generally focused on students’ perceptions of teaching rather than their actual learning. It may provide useful feedback on the teaching experience but if the aim is to improve attainment of learning outcomes, the focus should be factors that facilitate student learning (Denson et al, 2010). Similarly Nygaard and Belluigi et al (2011) suggests that when evaluation is based on de-contextualised learning, it ignores the learning processes of students, providing information on perceptions ultimately like a popularity test. This is supported by research by Frick et al (2010) who claim course evaluations traditionally used in higher education have few items that are empirically related to learning achievement.

Powney and Hall (1998) argue that student satisfaction questionnaires are counter-productive, partly because they are underpinned by an unproven assumption that satisfied students get better results. Concerns have also been raised by numerous researchers about what influences the results (see for example, Francis (2011) who cites studies showing students’ attitudes to a course before the start of the semester, prior interest and field of study all have an impact on ratings.) The impact that such monitoring initiatives have in actually improving the quality of teaching has also been questioned. Denson et al (2011) cite a study that found only 3–10% of instructors reported making major changes in their teaching as a result of course evaluations. However, some researchers do claim a positive relationship between course grades and student ratings of course satisfaction. Svanum and Aigner (2011) suggest this relationship is well established but not straight forward so that controversy remains about the interpretation and magnitude of this association, together with its implications. Additionally, Frick et al (2010) report on meta-analyses of studies that have examined the correlation of the relationship between items such as “This was an outstanding course” and student achievement, finding a moderate correlation. Nevertheless, whilst not doubting the existence of the relationship, Frick et al (2010) stress that it does not indicate how to improve teaching. As a whole these results imply that the reliability and validity of measurement tools require further examination. Similarly, the extent to which evaluation takes account of the strategic direction of the Business School, the factors that influence student attainment and the requirements of those paying the fees needs exploring.
Concerns about the prominence of student satisfaction measures relate to the ‘what’ of evaluation procedures in higher education, but criticisms have also been directed at the ‘when’ or timing of evaluation, and the ‘how’ evaluation studies are conducted. Firstly addressing the timing issues, formal procedures are often at the end of a module or course. Whilst there may be a clear rationale for such summative evaluation, the multiple purposes of evaluation, such as, learning (Easterby-Smith, 1994) may not be facilitated in this way. Nygaard and Bellugi (2011) suggest effective practice would involve conducting data collection from students at a time when reflection could positively impact or facilitate their learning. This supports views of a more strategic approach that incorporates evaluation throughout the learning cycle rather than solely at the end. Secondly, the standardized nature of processes utilized by many institutions may not reflect the diversity of subjects, students, stakeholders and learning outcomes of different disciplines. Construction of a tailored, course specific approach may yield more meaningful results. We suggest the ‘educational scorecard’ may provide the framework in which this could be developed.

3. Research methodology

An exploratory approach was adopted based on inductive reasoning. The researchers had a good appreciation of the breadth of issues from the literature review and their own experience. The subsequent primary research was aimed at engaging stakeholders in issues relating to the design of evaluation procedures. There were two key stages in the data collection. First; a benchmarking exercise which sought to review current evaluation practices in a number of comparable courses and meet three key objectives;

- draw on best practice from other institutions,
- share creative ideas and
- learn from their success

The second stage sought views of two stakeholder groups: students and employers to identify their requirements. Clearly, there are other stakeholder groups, such as the university quality department, government funders and the professional body; for the purpose of this study their requirements were represented by the formal regulatory and validation processes.

For both data collection exercises a non-probability sampling method was used. This relied on convenience or opportunity sampling, drawing on a sample of the researchers’ contacts. The purpose of the research was explained to participants and they were given the choice to be named in the report or remain anonymous. Data collection took place by telephone interviews, focus groups and questionnaires depending on the respondent group. To enable consistency for analysis, a structured questionnaire or feedback sheet was used for each group, covering the same key issues, informed by the literature review. In summary the respondents were:

- 5 course leaders from Post 1992 universities offering a similar professionally accredited postgraduate courses in the same discipline
- 14 students on such a course at Leeds Metropolitan University
- 4 employers who currently or had previously sponsored students on such courses

As an initial exploratory exercise, the research was intended to focus on the priorities above, which were often missing from existing research. The scope of this study did not allow a more comprehensive or representative sample to be surveyed. The researchers acknowledged this limitation but felt the nature of the investigation was such that any views from the defined population were equally valid to contribute to the existing body of knowledge. This limitation does mean that any findings will be tentative and may require more rigorous testing through a deductive approach before implementation. However as our findings suggest an adaptable model for specific institutions and courses, this may not be essential.

4. Findings

The findings provide an insight into the views of university and stakeholder group participants. They are specific to the university and course context under investigation and are not taken to be representative of the whole population as per the inductive approach described above. The research focussed around three key themes:

- Evaluation procedures undertaken
• The purpose of such evaluation procedures
• The involvement of different stakeholders in the process

These were subsequently analysed to give a composite view of evaluation of the course being investigated and so inform the design of our ‘educational scorecard’.

4.1 Evaluation procedures

Survey outcomes from each of the three groups of respondents (benchmark universities, students and employers) are addressed in turn below.

4.1.1 Benchmarked universities - the course leader perspective

Unsurprisingly, many universities surveyed undertook very similar processes but referred to them by different names. These can be categorised as:

• course surveys,
• meetings with students / student representatives,
• evaluation integrated into course content

Each of these categories will be examined in turn:

Course surveys: Responses from all of the institutions mention university or faculty developed approaches to surveying students. With regards course level surveys, for two institutions the surveys were centrally developed with little opportunity to influence content. The third took part in the HEA Post Graduate Taught Experience Survey (PTES) characterised by the respondent as “the postgraduate equivalent of the National Student Survey”. As with the institutional surveys the questions were generic but information could be broken down to course level. The reliability and validity of the surveys was unknown but still proved helpful, as one respondent said “The processes can provide a useful insight into issues that are important to students but it is difficult to assess how representative they are. It is also heavily focussed on how students feel rather than other potential measures”.

Meetings with students / student representatives:

Four of the 5 universities conducted meetings with entire course cohorts - in some cases as a supplement to the representative support that was also available. It was suggested by one participant that this does not always work, one university with a post graduate forum rated it as neutral as “only a few from the course partake in it”.

Three of the 5 respondents outlined evaluation processes where students’ opinions were represented by their peers (elective representatives) at formal meetings. One institution with a twice-yearly ‘board of studies’ approach suggested that the process was good but “depends upon who is chairing the board and the trail of actions”. The second had annual ‘Enhancement and Development Meetings’. A third institution held twice-yearly ‘staff/student liaison committees’ because those on the course responsible for the employment relations agenda “want them to learn about working with representatives and the consultative committee approach” this was not however a formal assessed component of course content, elements of which are addressed in the next section.

Before moving on, it is interesting to note that two of the three institutions that mentioned student representatives identified the outcomes of staff/student meetings and follow up on actions thereon, as key to their success.

Evaluation integrated into course content:

One university identified how they had integrated an evaluation process into their course content. The first institution required students to run an end of course forum attended by employers and teaching staff as part of an assessed skills development activity. The one-day forum began with a half day exploring how students had developed over the programme, this element was assessed and was attended by employers. The second half of the day required students to participate in a whole course evaluation event lead by the course leader and module team. Commenting on the latter evaluative
component the Course Leader said "This is a way to have feedback across modules to see what is working as a coherent whole".

4.1.2 The student perspective

In a review of existing procedures, students were generally aware of both informal and formal methods of evaluation and recognized the importance of evaluation procedures for course development. Although one student commented ‘There is no point in conducting an evaluation if the college is not going to do anything with the information provided. I think a good Learner Evaluation form tells the students how the information will be used and when and where the findings will be published.’

The range of evaluation procedures cited included: module evaluations, online discussion forums, verbal/email feedback, assignment/exam results, university league tables, student surveys, engagement in lectures and pass rates. One student with disabilities suggested measures of diversity and how we deal with students with disabilities or special requirements should be incorporated into evaluation procedures. Some concerns were expressed regarding the completion of in class evaluations in the presence of the tutor. One student suggested that despite a broad range of procedures existing, there may be a lack of triangulation in analysis of the information.

The value of learning was described by students in terms of transfer of learning to the work context, either in their day to day roles or in terms of career progression and development. However, there were mixed views about the University's role in the transfer process and whether this was the university's responsibility or that of the students themselves. For example, some students are in low level jobs and felt they did not have an opportunity to transfer learning at the moment but would hope to do so as they progressed in their career. They suggested this could be assessed through a longitudinal evaluation (12 months and 3 years later), looking at career progression after completion of the course and surveying their organizations/sponsors. Diverse opinions were also expressed regarding the challenges of measuring return on expectation. It was acknowledged that the potential multiplicity of expectations could be problematic, particularly in a course governed by professional body requirements and university constraints.

4.1.3 The employer perspective

The first noteworthy finding is that of the 4 employers surveyed, there was very little employer evaluation of the courses on which they were sending students. There appeared to be greater confidence that the universities' quality assurance processes were satisfactory and fit for purpose. However, this meant that opportunities to evaluate transfer of learning were potentially missed.

The response to what employers might look for in a postgraduate, professionally accredited, course provided wide-ranging answers including: cost, location, time and flexibility.

4.2 Purposes of evaluation procedures

5.1.1 Benchmarked universities

All respondents were asked to identify the purposes to which they put the outcome of their evaluation exercises. The most frequently mentioned outcomes were:

- To provide information to enhance the course
- To measure student satisfaction
- A formal requirement of the Quality process
- To provide feedback on methods of delivery
- To measure if expectations have been met

Other purposes identified in the evaluation literature, such as comparing staff and courses (for staff, current and future students), measuring success rates, learning achievement, learning transfer and to provide information for staff feedback, appraisals, evaluation of teaching and teaching research; were considered areas outside of the concern of a course leader. Respondents were however concerned to say that these factors did however have a role to play for other stakeholders in the institution.
Despite these similarities, there are different attitudes amongst the course leaders to these processes. For a number it is box ticking, NSS and regulator driven compliance from the centre which led to local concern about the richness and timeliness of standardized processes. "We can’t do the course level one now as we’re not supposed to overload them with questionnaires. So we’ve lost some of the richness of data". For others it is about having a very positive faculty attitude to quality and having the centre involved in a supportive rather than an adversarial capacity.

That is not to say that courses have had to simply labour under university, regulator or indeed professional body systems. There was some evidence that courses had supplemented the University procedures with customized processes that enhanced the quality and depth of feedback received. There are novel approaches which are mentioned below. It seems then that the issue is where courses try and replicate university systems so that students experience 'déjà-vu'. Where novel approaches to evaluation have been identified, it seems that there is no reluctance from university stakeholders to let them take place. Whether it is because they do not mind or do not know about them is a question that remains unanswered.

Feedback had led to developments on a number of the courses, but not usually as a consequence of the rather blunt university wide surveys. It is difficult to measure the extent of change that had taken place as a result of evaluation procedures. This is because all of these accredited courses had undertaken a professional body-driven review within the last year which has resulted in extensive course revisions.

4.2.1 The student perspective

Students proposed a number of concurrent purposes of the course evaluation. The most highly rated purposes were:

- To measure if learning objectives have been achieved
- To measure if expectations have been met
- To measure student satisfaction
- To provide feedback to staff

4.2.2 The employer perspective

When asked what universities should be aiming for from their evaluation procedures respondents all agreed that feedback to staff and achieving learning objectives were the primary reasons for course evaluation. Also featuring strongly were:

- links to teaching staff appraisal/reward
- measures of student satisfaction
- feedback on methods of delivery
- course comparison data
- provision of information to enable appropriate course selection
- meeting of professional body standards

This is interesting when compared with employers’ decision making on course choice. It seems that practicalities are at the forefront with cost and location mentioned more than once and success outcomes ‘pass rates’ only mentioned on a single occasion. Other aspects of learning that were mentioned were delivery methods and content coverage. It could be suggested both of these were also practical considerations. The former around release from work issues, the latter around coverage of current work issues within a course to reduce the costs of further external training.

In contrast to the suggested aims for university evaluations the in-company assessment of learning was often without formal structure/measurement. Reference was made to: Student feedback, Performance and development review (PDR) discussions, line manager meetings and ‘it’s important candidates have the right qualification for their role’. It is clear then that there is little formal course evaluation, more that feedback is derived from spontaneous complaints/praise and employee performance management discussions where progress towards the desired academic outcomes were checked rather than the quality of the learning experience.
The results imply a trust in universities and their evaluation processes to validate the training on which employers choose to send their staff. This trust may also extend to the professional body accreditation of the university award. Employers appeared confident that if the course met these quality standards it would be fit for purpose and deliver intended workplace outcomes. Aspects such as teaching and student learning, typically of ‘pedagogic concern’ are recognised as important for the University to evaluate, however, back in the workplace the findings indicate a more pragmatic approach, where the more ‘business concerns’ of financial information and becoming ‘qualified’ were prevalent.

It should be noted that employer respondents were sourced via a professional body networking event. It is therefore likely that such sponsors strongly identified with the professional accreditation of the university awards and in many cases would have undertaken a similarly accredited award themselves.

4.3 Involvement of different stakeholders

4.3.1 Benchmarked universities

Perhaps surprisingly, for vocational, professionally accredited courses, engagement with employers on a regular basis was very limited. One institution used a ‘research network’ of industry specific practitioners as guest speakers, a second engaged through ‘placements’ both of which arguably, are not about direct course evaluation. A third said employer involvement was every 5 years at periodic review. Of the final two, one is hoping to achieve employer engagement through the new forum, the other said that this level of engagement happened at a business school level not at the course level.

Two institutions also mentioned the professional body and external examiners as stakeholders of the process. It seems that the nature of the relationship, focused on development as well as compliance, was key to securing a positive approach to the evaluation process and its outcomes.

4.3.2 The student perspective

The respondents perceived students, teaching staff and employers as stakeholders but did not suggest any other stakeholder groups. Students were seen as the primary ‘customer’. The role of employers was less clear, they were seen as sponsors, additional customers or even irrelevant to the university as the employer/student relationship was separate to the student/university relationship. One student commented ‘many employers just pay for the course as part of their commitment to Learning and Development and really their expectations in terms of learning outcomes and results are not as high as those of the students’. This separateness was echoed in the relationship between student and parents (or other family sponsors).

4.3.3 The employer perspective

All respondents, except one agreed that students, employers and teaching staff should be involved, one respondent also suggested line managers as distinct from a training or HR department. The way universities should collect feedback varied only slightly with most suggesting questionnaires and surveys and some suggesting focus groups. There was a strong view from employers that they should be engaged in evaluation at the end of the course and up to a year after a student has graduated, echoing the student perspective in 3.1.2 above. Many suggestions relating to students and teaching staff reflect common practice across the universities surveyed for this research. (As cited in section 3.1.1)

All respondents agreed that students were the primary customer, two noted the employer and one cited parents. Parents could be extended to refer to all family members who have a stake in the student’s investment in time and energy and their ultimate academic success.

5. The way forward

The findings supported the general tenet of the importance of evaluation in HRD literature, with all participants recognising the requirement for, and value of, course wide evaluations. The benchmarking data demonstrated ways that formal university procedures, which risk losing "some of the richness of data", could be supplemented by specific course level procedures to provide greater depth of information. Where these were integrated with teaching on the course, they also overcame the problem of low response rates and limited student engagement with the process.
The range of evaluation methods cited by both students and course leaders suggests that lots of information is collected in both formal and informal ways. However, the extent to which data is utilized in a coordinated manner could be improved. Students appeared to assume a more integrated approach than actually occurred. In practice, the results of different evaluation techniques were not combined into a meaningful whole, leading to a piecemeal rather than strategic view of evaluation. This leads to a reactive approach where small changes are made based on how students feel rather than improvements linked to higher attainment, echoing the concerns of Denson et al (2010) above.

Recommendations from students, one priority area, who took part in the research, supported the development of more flexible data collection, such as online discussion forums to enable them to provide timely feedback at any stage of the course and respond to comments made by other students. However, this may lead to the creation of unrealistic expectations around immediate change. Additionally if students are unaware of university procedures and compliance requirements, another priority area, it could mean they feel their desires for immediate and significant change are being ignored rather than pursued through rather lengthy processes and channels. Also as reported in 3.1.1 not all students participate in such processes, which could make data collected invalid due to limited quantity or skewed by those students with the most to say.

It was generally supported by all respondents that evaluation should have a student centered focus. This would identify the levels of student satisfaction, if learning objectives have been achieved and whether expectations have been met. However the learning objectives may be understood differently by each stakeholder e.g. the student, employer or professional body. The complexity of different audiences, even within the same stakeholder group, means views may differ. Embracing this diversity through involving different stakeholders in the evaluation process design and implementation may lead to a more meaningful collection of data. Better dialogue between stakeholders at the ex-ante, interim and ex-post (Devins and Smith (2010)) stages, may help. It should be noted that collection of evaluation data is not the end of the evaluation process. The importance of action planning is essential to lead to continuous learning and improvement, by all stakeholders. Practical consideration needs to be made as to how all relevant stakeholders’ views can be captured to provide manageable data for courses and universities to analyse and implement improvements.

Our conclusion is that the course evaluation for the course investigated, should include the following:

- A strategic focus (linked to professional body and university strategy)
- Clearly stated and communicated purpose and outcomes
- Integration into the course as a whole not just completed at the end
- Inclusion of appropriate stakeholders
- Engagement of students, in order to provide greater depth of information, through customized course level procedures that enhance rather than duplicate the standardised university approaches
- Triangulation of data from different methods to ensure a more balanced basis for decision making
- Where possible, benchmarking against internal and external data

As mentioned previously in section 1, researchers have utilized the Balanced Scorecard as a framework for academic institutions. The studies generally report on the application of the scorecard at organizational level. Our proposal is that a scorecard tailored to the specific requirements of the course would enable us to address the features listed above. In order to achieve this we recommend the development of a course specific ‘Educational Scorecard’. The balanced scorecard provides a framework against which long term performance can be evaluated. It identified 4 perspectives; ‘financial’, ‘customers’ ‘internal business’ and ‘learning and growth’ (Kaplan and Norton 1996). When applying the scorecard, individual Key Performance Indicators are captured under one of the four headings and these are measured and managed regularly, sometimes as part of a management dashboard. Translating this into a university environment is not simple. Currently, as reported above there is a lack of an integrated approach between relevant stakeholders but if universities are to survive in the increasingly competitive marketplace where the views of such stakeholders are arguably critical, then accurate measurement and timely responses may be the key to gaining high student satisfaction scores and other success measures.
Below you will find our proposed Educational Scorecard for the course investigated for this paper. As with Kaplan and Norton (1996) there are 4 areas which map the priority areas for course evaluation and management. Each perspective contains some examples of content but the lists are not exhaustive.

The Educational Scorecard contributes to the solution.

![Educational Scorecard Diagram]

**Figure 1**: The educational scorecard

Derivation of the Model from our Research

An analysis of the literature and our primary research shows information is gathered from qualitative and quantitative sources and for different purposes. Although student-centred-ness is a key driver for much of what is done and justifiably so, our collective evidence shows that many other factors contribute to a rounded, multi-dimensional evaluation. This will ultimately be of greater value and impact. The four quadrants of our Educational Scorecard summarise the areas we identify as key to this holistic approach. The upper quadrants are derived from data extracted from or needed by the key stakeholders; regulatory bodies such as the university and professional body and students. The lower quadrants represent the key purposes of evaluation in this context; to improve teaching and extend learning and growth. Our model seeks to capture the diverse and wide-ranging sources of data for evaluation identified earlier in our research. This analysis can be tested working in reverse and mapping each quadrant against the evidence here presented. It is evident that all quadrants of the model are discussed almost in equal measure in our findings. Therefore this is presented as a balanced Educational Scorecard. However in its application, we caution that it might identify specific courses where one quadrant is currently weighted (used) more in decision-making than the others. In which case, evaluation of the data collected should be interpreted and if necessary more data collected against the other quadrants to capture a more rounded picture from which to implement more informed decisions.

6. Application and significance to a wider audience

The basis of our study stemmed from a course leadership team inundated with evaluation procedures and sources of feedback, but concerned about the fragmentation, validity and reliability of this information. Consequently, we were questioning our ability to make effective and strategic improvements to the course based on information that was piecemeal and perhaps unrepresentative of the views of all of our stakeholders. The educational scorecard has given us a more balanced view!
We are still at an early stage of implementation but would claim the following initial gains:

- It gives a clearer perspective to the formal University and professional body evaluation procedures.
- It has made us integrate evaluation rather than complete it predominantly at the end. We begin discussing student expectations at the interview stage and again at induction so that we have jointly agreed expectations that we can review throughout the course.
- It has enabled us to identify areas where we need to obtain more input from stakeholders, for example, employer and sponsor feedback.
- Pedagogic evaluation has led us to develop an induction framework for associate lecturers.
- It helps us demonstrate our success. The articulation of our research outputs and development activities establishes credibility for the teaching team, hopefully leading to greater trust and student engagement.
- It establishes a framework for comparative analysis of the course over time.

This study has focused upon one course in one university, drawing on data from some of the relevant stakeholders and comparable courses in other universities. Its findings were not meant to be automatically applied more widely to different courses in other institutions. However, the Educational Scorecard is designed to be adaptable, with scope to vary the titles and content of each quadrant and we would therefore argue that it has a much broader relevance than just this course. In the current environment of increased scrutiny of university performance and the publication of league tables it is imperative that course leaders have appropriate management information to demonstrate success and drive improvements. The significance of league tables has been questioned as they combine a broad range of measures which potentially conceals specific areas of strength or weakness. The lack of universally agreed criteria also leads to variance in results; there is an inconsistency between different UK lists and also a disparity with global rankings. We would contend that a meaningful collection of data at course level is a more appropriate means of strategic performance monitoring. By modifying the measures in each quadrant to reflect the priorities of the particular course, university and professional body, the methodology of the educational scorecard could be applied to a wide range of different courses.

7. Further research

Future adopters of the Educational scorecard will need to assess its applicability beyond the Master's level course considered herein, initially by replicating its usage in similar degrees at other institutions. Following, there is a need to understand its ability to provide appropriate measurement across different subject areas, and also at different levels of university study - for example at undergraduate level. Adoption at undergraduate study would necessitate the consideration of other appropriate agendas and stakeholders. The researchers here suggest that employability could become a factor that is addressed as an alternative to direct employer/sponsor input. This however needs further evaluation.

It may also be that a similar approach to the one outlined on this paper could be adopted within secondary education but without further investigation at this stage it is inappropriate to make such an assertion.

Above all there needs to be longer term research to establish the longevity and impact of activities and proposals such as this one. Does it offer a sustained qualitative difference in evaluation and outcomes or is it a novel approach which becomes increasingly sidelined as yet another evaluation activity?

References


Research Philosophy Debates and Classifications: Students’ Dilemma

Marcia Mkansi and Edwin Asiamah Acheampong
University of Bolton, UK
mm1lia@bolton.ac.uk
eea1lis@bolton.ac.uk

Abstract: Research philosophy classifications such as ontology, epistemology, and anxiology and their conflicting applications to the ‘quantitative-qualitative’ debates, are a major source of dilemma to research students in establishing their relevance to subjects areas and discipline. A number of studies have used different descriptions, categorisations and classifications of research paradigms and philosophies in relation to research methods with overlapping emphasis and meanings. This has not only resulted in tautological confusion of what is rooted where, and according to whom; but raises a critical question of whether these opposing views are enriching knowledge or subtly becoming toxic in the field? . This paper puts forth a student voice towards these debates and aims to provoke research advocates from their peripheral standpoint to become concerned about this subtle but deepening concern of students and their future impacts. A concerted effort in this direction should eventually result in the development of a planned, systematic framework and procedure that show some consensus to bail research students from these bewildering classifications and debates. The paper briefly reviews, discusses, and analyses these research philosophy classifications and debates and provides a mapping thereby through literature. Then, assesses how they impact on research students through case studies based on three North West Universities in the UK. Responses were elicited using structured interview questionnaires where students fall into different faculties and subject groupings for comparison purposes. Although the findings paint a grim picture of research, they are not conclusive to all UK students as the sample studied is skewed geographically. Future studies must survey the impact from other geographical locations. It is the conglomeration of these studies that will provide the ‘real’ magnitude of the impact on research students. This paper contributes to discussions on research methods and calls for a consensus in the field of research.

Keywords: research philosophy, debates, students, dilemma

1. Introduction

The difficulty in conducting research today is heightened by the incoherent classification of research philosophies such as epistemology, ontology, anxiology and doxology and the quantitative-qualitative dichotomy debates, in a way that those who made it are unlikely to be affected by it. A number of studies (Saunders et al., 2009; Ritchie and Lewis, 2003; Guba, 1990; Guba and Lincoln, 1989) have used different descriptions, categorisations and classifications of research paradigms and philosophies in relation to research methods with overlapping emphasis and meanings. This has resulted in tautological dilemmatic confusion of what is rooted where, and according to whom – doubt as to how it informs future studies and potentially negative impact on those that are subject to their application especially, research students. The mounting debates have taken a distinctive turn that causes dilemma to research students in establishing its relevance to subject areas and disciplines. This paper briefly reviews, discusses and analyses the debates and classifications by way of literature reviewed. The dilemma is examined in this research through a series of case studies conducted with research students from three different North West Universities in the UK. The students are divided into different faculties and subjects groups to allow for comparison and to obtain different perspectives and understanding of these impacts. Carefully structured interview questionnaires are used to collect the data. The wide-range of questions employed enabled different facets of the debates to be assessed. The overarching aim is to find out whether and how these research philosophy classifications and debates impact PhD students’ research work, such as their choice of research approach, and provide a mapping of these classifications, on one hand, and assess the implication for the future of research.

2. Research philosophical stance debates

Proponents of research philosophies (Saunders et al., 2009; Guba and Lincoln, 1994; Becker, 1996) have engaged and displayed their knowledge and beliefs in what appeared and got interpreted as paradigm “wars”. Although their definitions of ontology, epistemology, and anxiology have a common
theme with a bit of different meaning and emphasis; there seem to be no consensus in the classification and categorisation of these paradigms. A thorough literature review (Saunders et al., 2009; Will et al., 1997) of these philosophies and research methods has overlapping evolutionary process. The evidence ranges from the original presentation, description and categorisation outlined amongst other major philosophical advocates (Burrell and Morgan, 1979; Guba and Lincoln, 1989; Guba, 1990) to recent philosophical scholars (Saunders et al., 2009; Ritchie and Lewis, 2003). For example, Ritchie and Lewis’s (2003) description and classification of ontological and epistemological stances is different from that of Saunders et al. (2009). Ritchie and Lewis’s (2003) ontological perspective include realism; materialism, critical realism, idealism and relativism; and the epistemological perspective include positivism and interpretivism. Hence, Saunders et al.’s (2009) and Guba and Lincoln (1994) indicate a perspective that views philosophies (i.e. positivism, realism, interpretivism, and pragmatism) from an ontological, epistemological, axiological stance. An even overlapping classification of these philosophies is that of Guba and Lincoln (1989); which links positivism, post–positivist, and constructivist to critical realism. A further example in the context of philosophy is pragmatism which, is thought to have evolved from realism described as Peirce realism by Thayer (1981). These philosophies are not entirely different. Put differently, they all share a common set of assumptions, and their commonalities identify these philosophies as examples of broader philosophies. However, whilst they share critical assumptions, they emphasize very different implications of those assumptions. And while they all focus on explaining methodological differences in research, they adopt different categorisation and classification. Given these differences, it is not surprising that these philosophical debates have generated a dilemma for research students. Indeed, recognising the potentially adverse impact on research students should call for a concerted effort to standardize the philosophies.

3. Philosophy and research approach debates

There is much ongoing debate on where a particular method (i.e. qualitative, quantitative or mixed method) is rooted or founded in relation to these philosophies (Johnson and Duberley, 2000; Bryman, 1984; Morgan and Smircich, 1980; Caelli et al., 2003). Even Guba and Lincoln (1994) acknowledged the ongoing patents of paradigm to research approach. Few exemplary discussions of these debates are provided on table 1 below:

**Table 1:** Philosophical debates for: Mixed methods, qualitative and quantitative approach

<table>
<thead>
<tr>
<th>Research approaches</th>
<th>Philosophical debates for the approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed Methods</td>
<td>Mixed method is believed to be rooted in pragmatism (Denscombe, 2007; Johnson and Onwuegbuzie, 2004); but Barrett (2010) asserts the opposite and considers critical realism to be the theoretical foundation of mixed method research.</td>
</tr>
<tr>
<td>Qualitative approach</td>
<td>Guba and Lincoln (1994) identified four paradigms that compete in qualitative research, namely, positivism, post-positivism, critical theory and constructivism. However, other scholars (McNabb, 2008; Denzin and Lincoln, 2005) assert that interpretive and critical paradigms are central to qualitative approach. A detailed analysis of these debates is fully discussed amongst other scholars (Guba and Lincoln, 1994; Denzin and Lincoln, 2005).</td>
</tr>
<tr>
<td>Quantitative approach</td>
<td>Scholars (Polit and Back, 2008; Steen and Roberts, 2011) assert that positivist and naturalist are philosophies for quantitative approach. Furthermore, Alvesson and Skoldberg (2009) add post-positivism, social constructionism, and critical realism as other philosophical stances.</td>
</tr>
<tr>
<td>Qualitative vs. Quantitative approach</td>
<td>The debates stretch further from which different philosophies is best for a single research approach (i.e. pragmatism against critical realism for mixed methods); to whether a particular philosophy is for qualitative or quantitative approach (i.e. positivism for qualitative vis-à-vis quantitative). This is best demonstrated in research findings by amongst other scholars (Bryman, 1984; Becker, 1996).</td>
</tr>
</tbody>
</table>
3.1 Philosophy debates for neither of the approaches

Whilst the scholars battle out on the foundation of philosophies or paradigms to research approach, Johnson and Onwuegbuzie (2004) argue that the differences in epistemological beliefs and logic of justification do not dictate what specific data collection and analytical methods to utilise, therefore, should not prevent the exploitation of any methods. In support of this, Guba and Lincoln (1994) stress that the rooting of approaches is secondary to paradigm and do no limit the use of either qualitative or quantitative from any research paradigm. Differences of trivial nature paid to philosophies in relation to research approach are greatly discussed by Becker (1996).

4. Philosophy and field of study debates

Then follows the debate on which philosophy is for social sciences or natural sciences? The survey findings by Orlikowski and Baroundi (1991) and Alavi and Carlson (1992), for example, report that the philosophies found to be widely applied and more popular in Information systems was positivism, and links the philosophy to the field of social sciences (Steinmetz, 2006). But, Polit and Beck (2008) associate positivism to natural sciences. In a different view, Bryman (1984) associates positivism to social research that applies natural science. This leads to subject rooting to different fields of study, for example, Information system’s roots is found to be overlapping with different field of studies such as social sciences, computer science, and business studies (Orlikowski and Baroundi, 1991; Hirschheim, 1985; Steinmetz, 2006). Hence the debate boarders on whether quantitative is for natural sciences (Polit and Beck, 2008; Steen and Roberts, 2011), and the case of qualitative approach in social sciences as is widely prescribed by Babbie (2007).

5. Philosophy and subject debates

The philosophical foundation debate gave rise to contradicting arguments as to which philosophy is best for a particular subject. For example, three different philosophical views have been identified for information systems (IS), namely, critical social theory (Ngwenyama and Lee, 1997; Orlikowski and Baroundi, 1991). Pragmatism (Agerfalk, 2010; Goldkuhl, 2008); critical realism (Hjorland, 1998; Dobson, 2002). Yet, the survey findings by Orlikowski and Baroundi (1991) and Alavi and Carlson (1992) indicate that the philosophy found to be more popular in Information systems is positivism.

5.1 Pragmatist view to the subject of information system

Goles and Hirschheim (2000) brought pragmatism into Information systems. The importance of pragmatism to information systems has been acknowledged amongst others by Agerfalk (2010); and Baskerville and Myers (2004). The emphasis by these scholars is that information systems is often seen as pragmatic discipline with a prominence on practical research, theory and practical implications.

5.2 Critical social theory view to the subject of information system

The connection between CST and IS had been launched from two fronts: firstly, relationship between science, theory and practice; and the social action and meticulous knowledge in which it is based (Alvesson and Willmott, 1992). Klein and Hirschheim (1993) emphasised the relationship of pragmatism and IS to work, social and interaction. Examples of CST application to IS research include, amongst others, studies by Alstyne and Brynjolfsson (2005); Orlikowski and Baroudi (1991); and Basden (2002).

5.3 Critical realism view to the subject of information system

Several scholars (Scott, 2007; Carlsson, 1989; Barrett, 2010; Dobson, 2002) have endorsed CR as an epistemological stance for information systems. The emphasis is towards its ability to address natural and social sciences, which offer a platform for use of variety of methods. Hence, its ability to take a realist stance by pointing out the limitations of positivism and interpretivism individually, and subsumes critical social theory (Mingers, 2004; Wikgren, 2004).

6. Research terminology debates

Advocates of research methods (Srivastava and Rego, 2011; Saunders et al., 2009; Khotari, 2006) have used different terminologies that are contradictory one to another, which leave students staggering as to which is which, and why there is no consensus in the scholarly fields. Amazingly, even Johnson and Onwuegbuzie (2004) picked on some of this confusion; which was on the
reference or treatment of epistemology and method as synonyms. Yet, not much thought is given to how it impacts on students, who widely apply and rely on these scholars’ guidance for research directions and clarifications.

6.1 Examples of terminologies on research approaches

A number of studies have used different descriptions of the main research approaches with common themes, categorisations and overlapping emphasis, which is, in fact, qualitative and quantitative. For example, quantitative is also known as empirical research (Hinchey, 2008); deductive, explanatory (Saunders et al., 2009; Engel and Schutt, 2005). On the other hand, qualitative is also known as exploratory research (Neelankavil, 2007; Engel and Schutt, 2005; Bernard, 2006); inductive (Saunders et al., 2009; Engel and Schutt, 2005); and formulative (Khotari, 2006). Whilst the latter scholars acknowledge the similarity of these approaches, some other scholars treat these approaches separately. For example, Hinchey (2008) clearly indicates that empirical research is quantitative, but, Srivastava and Rego (2011) described empirical and quantitative as separate types of research approaches. In a thorough review of these terminologies, it appears rather, to have amounted to a tautological confusion of these different descriptions of research approaches. This is because all other types of research approaches are variations of either one or combination of qualitative, quantitative, and the combination of the two often referred to as mixed method or multi methods depending on the, field of study (i.e. social science, natural science), or the source of scholarly reference (i.e. Saunders et al., 2009; Khotari, 2006, etc). In support of the fact that there are only three distinct approaches to research is Bryman (2006).

6.2 Examples of terminologies on research approach, techniques and strategies


7. Research design and methodology

This research adopted a qualitative approach and considered case study as strategy. This was mainly due to the strength of case study in answering ‘how’ questions and providing in-depth understanding of phenomena as widely described by Yin (2003). Structured interviews were considered as appropriate technique for extracting comparable findings, as echoed by Khotary (2006). The study focused on three North West Universities as case studies to provide in-depth perspectives of how research philosophical debates and classifications impact on PhD research students’ understanding and decision of research approach. The aim was addressed by:

- Reviewing literature in order to map different philosophical debates by different scholars, from which the structured questions for interview purposes were derived.
- Purposively sampling PhD research students from different subjects, faculties, and Universities, and used both personal and telephone methods of interviews.

The purpose of focusing on three different University, PhD research students, diverse subjects, and varied faculties was to provide comparable multi-perspectives of impacts to understanding and their decision of research approach. Saunders et al. (2009) assert that this kind of approach is good for establishing whether the findings of one case occur in another case. At best, the limit to three Universities and the overall approach helped to eliminate some of the fervent criticisms against case study relating to control, access, time, and travel and analysis cost as expressed by among other scholars (Gable, 1994; Rowley, 2002; Darke et al., 1988). The selected cases in the North West are: The University of Manchester, University of Salford, and the University of Bolton. The usefulness of studying a limited number of cases in-depth have been discussed and expressed by John and Onwuegbuzie (2004).

7.1 Case analysis

The study collected at first data that aimed at providing profiles of 26 PhD students that were purposively sampled, from which comparison can be made for each faculty. The purpose for profiling was to gain insight into the extent of understanding of research philosophical debates and classifications by each PhD students and to find clues as to how the students decide on research
Marcia Mkansi and Edwin Asiamah Acheampong

approach. This aided in providing evidence and determining similarities and differences in levels of understanding and decisions on research approaches. The total number of respondents from Bolton University was 16 (61.5%), from which 11 were in their final year, 4 in the second year, and 1 a first year student. Manchester University’s respondents comprised of 7 (26.9%) students, from which 6 were in third year and 1 second year PhD students. Hence, Salford University participants were 3 (11.5%), of which 1 is on final year and two on second year. Respondents came from 61.5% male and 38.5% female students from diverse subjects groups which comprised of: business, science, engineering, music, art and theology. The breakdown of students faculties are: 46.2% social sciences, 19.2% humanities, 23.1% science, and 11.5 engineering (table 2). Data was coded and analysed through SPSS which allowed for some quantification, and mapping of different approaches chosen by PhD students in each faculty, and subjects; whilst enabling a point of in-depth comparison and contrast of impacts to understanding and decision making.

Table 2: Faculty distribution of respondents

<table>
<thead>
<tr>
<th>Students’ faculties</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Sciences</td>
<td>46.2</td>
<td>46.2</td>
<td>46.2</td>
</tr>
<tr>
<td>Humanities</td>
<td>19.2</td>
<td>19.2</td>
<td>65.4</td>
</tr>
<tr>
<td>Sciences</td>
<td>23.1</td>
<td>23.1</td>
<td>88.5</td>
</tr>
<tr>
<td>Engineering</td>
<td>11.5</td>
<td>11.5</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

7.2 Case results: PhD students from Bolton and Manchester

The interview canvassed opinions on research philosophical debates in relation to its varied classifications, research approach, field of study, subject, and research terminologies debates.

7.2.1 Philosophical stance debates findings

This part of the study highlights key concerns, which PhD students’ have towards research philosophy debates and its classifications. Of particular interest were the perceptions of understanding and relevance of research philosophies in choosing a research method. The findings indicate similar patterns of confusion in the classification and understanding of research philosophies by PhD students across different faculties and year of study. Whilst 61.5% of students indicated awareness of different classifications and debates surrounding research philosophies 7.7% of respondents mainly science and engineering students have never heard of the ontology, epistemology, and anxiology terms, and yet were on final year of PhD. The other 30.8% split of years in study and faculties were not aware of the classifications, though mindful of ‘qualitative-quantitative’ debates. Majority of respondents, 69.2%, reported that it was confusing and difficult to understand the debates and its classifications (see table 3). This came as a sharp contrast from 11.5% that thought research philosophies were easy and straight forward to understand. The students’ grounds for difficulty and confusion include: ambiguity, different descriptions, and categorisation of same philosophy. Could this be confirming Silverman’s (2010) dangerous view of research dichotomies? What does the future role as researcher and PhD supervisors entails for the 7.7% students on final year, and yet not aware of ontology, epistemology, and anxiology terms?

A disconcerting finding is the fact that 23.1% research students from science and engineering faculties (majority of respondents from the faculties) were less concerned and did not have much knowledgeable on the philosophical debates and classifications, and felt it was less crucial in their studies. This perception is more inclined to be in dis-accordant with the rest of other faculties’ students (76.9%), who indicated a disentangled view describing them variously as, necessary but irrelevance vis-à-vis relevant but unnecessary, or necessary and relevant. The views of science students combined with those that were thorn in between unnecessary and relevance is in accordance with Becker’s (1996) unnecessary view of differences paid to philosophies in relation to research approach. Of greater interest however, was the students’ rationale for associating with research philosophies. Only 7.7% suggest that their inclination to philosophy is more of personal belief; with some 38.5% indicating that their inclination is influenced by qualitative and quantitative approaches association to a philosophy. The significant mix of students (19.2%) from science, humanities, and social science (excluding the 7.7% that never heard of the terms), claim to have been influenced by their supervisors towards a particular philosophy. Although 7.7% indicated a kind of negative enforcement towards the philosophies, it will be highly risky to assert any positive or
negative influence by supervisors. Therefore, the question of supervisors’ influence towards a particular philosophy remains a research question. Whereas, 30.8% claim to be influenced by a combination of self belief and research approach, a small number, 3.8%, from social science claim to be influenced by field of study.

The students’ primary view of research approaches to philosophies creates intellectual dissonance, that is, in contradictory belief to Guba and Lincoln’s (1994) secondary view of approaches rooted in philosophies. Interestingly, the 30.8% students who assert a combination of self belief and research approach as their motive towards research philosophy have no firm philosophical stance. Put differently, their research philosophy changes with the research approach and problem. Clearly, there is a huge disparity in research philosophies as viewed and understood by students sampled and that described by the majority of research philosophy advocates (Saunders et al., 2009; Ritchie and Lewis, 2003; Polit and Beck, 2008).

Table 3: Comprehension of classification and debates

<table>
<thead>
<tr>
<th>Comprehension of classifications and debates</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very straight forward</td>
<td>3</td>
<td>11.5</td>
<td>11.5</td>
<td>11.5</td>
</tr>
<tr>
<td>Confusing</td>
<td>14</td>
<td>53.8</td>
<td>53.8</td>
<td>65.4</td>
</tr>
<tr>
<td>Difficult to understand</td>
<td>4</td>
<td>15.4</td>
<td>15.4</td>
<td>80.8</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>7.7</td>
<td>7.7</td>
<td>88.5</td>
</tr>
<tr>
<td>No Response</td>
<td>2</td>
<td>7.7</td>
<td>7.7</td>
<td>96.2</td>
</tr>
<tr>
<td>99.00</td>
<td>1</td>
<td>3.8</td>
<td>3.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

7.2.2 Philosophy vs. research approach debates findings

A little over 30% of respondents indicated that qualitative approach can be viewed from any philosophical stance (table 4). Of the number of respondents (53.8%) that associated the qualitative research approach to some kind of philosophy, the breakdown is as follows: interpretivism – 19.2%, critical realism – 11.5%, realism – 3.8%, positivism – 15.4%, and critical social theory – 3.8%. The majority, 19.2%, sides with the position of a number of existing literatures (McNabb, 2008; Denzin and Lincoln, 2005).

Table 4: Philosophical stance for qualitative approach

<table>
<thead>
<tr>
<th>Phil. stance for qualitative</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpretivist</td>
<td>5</td>
<td>19.2</td>
<td>19.2</td>
<td>19.2</td>
</tr>
<tr>
<td>Positivist</td>
<td>4</td>
<td>15.4</td>
<td>15.4</td>
<td>34.6</td>
</tr>
<tr>
<td>Critical Realist</td>
<td>3</td>
<td>11.5</td>
<td>11.5</td>
<td>46.2</td>
</tr>
<tr>
<td>Realist</td>
<td>1</td>
<td>3.8</td>
<td>3.8</td>
<td>50.0</td>
</tr>
<tr>
<td>Critical Social Theorist</td>
<td>1</td>
<td>3.8</td>
<td>3.8</td>
<td>53.8</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
<td>30.8</td>
<td>30.8</td>
<td>84.6</td>
</tr>
<tr>
<td>No response</td>
<td>3</td>
<td>11.5</td>
<td>11.5</td>
<td>96.2</td>
</tr>
<tr>
<td>Never heard of the terms</td>
<td>1</td>
<td>3.8</td>
<td>3.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The apparent lack of awareness of the relationship between the research approaches and the underlying philosophies surface in the cases for quantitative and mixed methods. (see table 5 and 6 for breakdowns). In summary, the majority of students could not link a particular approach to a particular philosophy. There seem to be an inclination towards books that are in similar view to Johnson and Onwuegbuzie (2004); Guba and Lincoln (1994), who do not limit the use of approaches from any philosophical stance. The rationale for their views of philosophies varies from topic under study, research approach to research books for the majority of respondents (30.8%). Since majority of respondents indicated awareness of different classifications and debates, but their views of philosophies leaned towards books in line with Guba and Lincoln (1994) and Johnson and Onwuegbuzie (2004), would this imply that students found less confusion and more flexibility in the non attachment of philosophies to research approaches? Although there are patterns of confusion in understanding philosophies, there was almost universal agreement that research problems inform the research approach by most students (50%). Hence, the other 46.2% split between philosophical
stance – 15.4%; faculty preference – 15.4%; combination of approach, subject, and philosophy – 15.4%. The former is consistent with the majority of research methods advocates (Sauders et al., 2009; Khotari, 2006; Polit and Beck, 2008). Research students disagree, in part, with the manner in which the advocates of each philosophy put them forward. While they are supposed to be complementary, research students find them incoherent and are left in a dilemma, because there is lack of consensus among the writers of the research literature.

**Table 5: Philosophical stances for quantitative approach**

<table>
<thead>
<tr>
<th>Phil. stance for quantitative</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretivist</td>
<td>4</td>
<td>15.4%</td>
<td>15.4%</td>
<td>15.4%</td>
</tr>
<tr>
<td>Positivist</td>
<td>2</td>
<td>7.7%</td>
<td>7.7%</td>
<td>23.1%</td>
</tr>
<tr>
<td>Critical Realist</td>
<td>4</td>
<td>15.4%</td>
<td>15.4%</td>
<td>38.5%</td>
</tr>
<tr>
<td>Other</td>
<td>12</td>
<td>46.2%</td>
<td>46.2%</td>
<td>84.6%</td>
</tr>
<tr>
<td>No response</td>
<td>3</td>
<td>11.5%</td>
<td>11.5%</td>
<td>96.2%</td>
</tr>
<tr>
<td>Never heard of the terms</td>
<td>1</td>
<td>3.8%</td>
<td>3.8%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>100.0%</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6: Philosophical stances for mixed methods**

<table>
<thead>
<tr>
<th>Phil. stance for mixed method</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical Realist</td>
<td>3</td>
<td>11.5%</td>
<td>11.5%</td>
<td>11.5%</td>
</tr>
<tr>
<td>Realist</td>
<td>4</td>
<td>15.4%</td>
<td>15.4%</td>
<td>26.9%</td>
</tr>
<tr>
<td>Other</td>
<td>13</td>
<td>50.0%</td>
<td>50.0%</td>
<td>76.9%</td>
</tr>
<tr>
<td>No response</td>
<td>5</td>
<td>19.2%</td>
<td>19.2%</td>
<td>96.2%</td>
</tr>
<tr>
<td>Never heard of the terms</td>
<td>1</td>
<td>3.8%</td>
<td>3.8%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>100.0%</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

### 7.2.3 Philosophical debates vs. field and subject of study findings

Although 38.5% students indicate awareness of different philosophies endorsed for their field of study and subject areas, the majority 57.7% attest no knowledge paid to field of study and subjects, rather limit their awareness of differences to ‘qualitative-quantitative’ debate. There was minor uniform connection of a single philosophy to field or subject area since the majority had limited knowledge. Those that were aware, claim to have justified their selection of one over another by associating the particular philosophy to research problem, values, and supervisors. However, noting the difficulty and confusing in navigation through the philosophies, labelling it a grey area. For example, out of the 34.6% science and engineering students, only 3.8% linked quantitative to positivism. This is in sharp contrast to the gospel preached by advocates of positivism (Polit and Beck, 2008; Alvesson and Skoldberg, 2009); who assert that positivism is the philosophical stance for science and engineering. It further confirms earlier reports by these students that philosophical stance was less relevance to their research. Similarly, only 19.2% linked qualitative to interpretivism out of the 46.2%; another contrast to advocates of social sciences (McNabb, 2008; Denzin and Lincoln, 2005). An arising question is whether philosophical stance is not drummed much to science and engineering students in comparison to other faculties? Or the gospel preached by advocates of: positivism to science and engineering; and interpretivism to social science is not heard or applied? Is it becoming more less important to students?

### 7.2.4 Research terminologies debates findings

The majority of students’ 65.4% description of research terminologies seems to be in favour of books that describe research terminologies similarly to Saunders et al. (2009), as compared to those that use the terms interchangeable like Newman and Benz (1998). However, a significant minority 34.6% give a disconnection that is neither right nor wrong subject to references, between research proponents’ and students, the impact of which is directly related to the proponents’ different descriptions of same research terminologies. In the final view, 84.6% students’ calls for consensus, training and clarity that would facilitate diversity between students’ and proponents’ of research, with 15.4% questioning philosophical value and relation to their studies.
8. Conclusions and recommendations

The majority of PhD students, regardless of their year of study, subject and faculty, held a universal perception of confusion and difficulty in understanding research philosophical debates and classifications. The fact that almost every PhD students interviewed showed some form of dilemma towards research philosophical debates and classification, beckons a need for consensus by philosophical advocates. The contradictory perception between research students and philosophical advocates symbolize an alarming disparity, which research institutions and scholars, for the credibility of research, must consider taking decisive actions. This study has put forth a thought provoking discussion pertaining to the impacts of the debates on students’ understanding and choice of approach, and encapsulates the debates as evidenced in the literature reviewed. However, the onus remains with research methods and paradigms’ advocates to develop frameworks and devices that would bail research students from the dilemma, and rather provide systems that help students’ get a grasp of a difficult field. In reflection of the findings moving forward, the question is for proponents of research philosophies and methods to assess, whether the difference are enriching students’ knowledge as intended? Or becoming toxic to the field of research and minds of young researchers? If enriching the knowledge, how come the majority of students’ find it difficult and confusing? And if the confusion suggests some slow poison to the field of research, how is this going to impact the future generation? Especially where, final PhD students have not heard of the terms and those with knowledge of the terms find it hard? How does it impact supervision of prospective PhD students?

References

Marcia Mkansi and Edwin Asiamah Acheampong

Eating our own Cooking: Toward a More Rigorous Design Science of Research Methods

John Venable and Richard Baskerville

1 Curtin University of Technology, Perth, Western Australia, Australia
2 Georgia State University, Atlanta, Georgia, USA

j.venable@curtin.edu.au
baskerville@acm.org

Abstract: This paper argues that Design Science is an appropriate paradigm for research into Research Methods. Research Methods (along with their tools and techniques) are purposeful artefacts, designed and created by people to achieve a specific purpose – i.e. to create new, truthful knowledge. Like other artefacts, research methods vary in their fitness to purpose, i.e. in their utility, depending on their fit and appropriate application to the particular purpose, contexts, and contingencies for which they were developed. Design Science Research aims at developing new purposeful artefacts with evidence of their utility. Applying a DSR perspective to research methods should yield increased utility in the application of research methods, better guidance in applying them and greater confidence in achieving the desired outcomes of applying them. Based on these premises, this paper reviews the basic concerns and issues in Design Science Research (using the balanced scorecard as an example purposeful artefact), then analyses the logical consequences of taking a Design Science perspective on research methods (using the Partial Least Square approach as an example research method purposeful artefact). First, it analyses the various purposes of research methods to clarify the alternative and competing design goals of research methods. Second, it analyses and characterises the types of purposeful (design) artefacts that comprise research methods. Third, it considers issues of the evaluation of research methods. Fourth and finally, it considered the development of design theories of research methods.

Keywords: research method, research design, design science research, evaluation, design theory, research rigour

1. Introduction

There is a long history of discussion, debate, and publication regarding research methods (e.g. 1961; Campbell and Stanley 1963; Bunge 1967; Glaser and Strauss 1967; Miles and Huberman 1994). The ideas and methods used in research have developed and evolved gradually, especially as the research domains and purposes have evolved. The discipline of Information Systems is a recent participant in this debate owing to its comparatively recent origin as a field. However, the domain of information systems has presented methodological challenges that have led to further progress in research methods to better and more fully address the IS domain (e.g. Mumford, Hirschheim et al. 1985; Galliers and Land 1987; Benbasat 1991; Mason 1991).

The development and evolution of research methods has often proceeded in an ad hoc and non-rigorous process, because scientific communities are social phenomena with their own social institutions and mores (Berger and Luckmann 1966; Kuhn 1970; Latour 1987). New research methods and guidelines are posited and developed, written up and communicated to the research community, generally as research essays. But key aspects of rigour in their development may be lacking, such as weak attention to their intended goals and purposes, measures of their effectiveness, efficiency, or efficacy, evidence that the research method meets such goals or qualities, and comparisons to existing methods that provide evidence that the method better achieves its goals.

Because research methods are designed artefacts, they could be better developed using rigorous design science research (DSR) approaches. A DSR approach to the design and development of new or improved research methods would provide a more rational and consistent basis for assuring both their utility and rigour. The rigour from using a DSR approach would come primarily from (a) more precisely stated design theories about research methods and (b) more rigorous evaluation of research methods. Explicating such elements may even help make the peer reviewing of the resulting research more predictable and less capricious.

A seminal paper on DSR by March and Smith (1995) provides support for the use of Design Science Research for the development of research methodology.
“Natural science uses but does not produce methods. Design science creates the methodological tools that natural scientists use. Research methodologies prescribe appropriate ways to gather and analyze evidence to support (or refute) a posited theory. They are human-created artifacts that have value insofar as they address this task.” (p. 258)

While March and Smith do not assert that design science creates the methodological tools that behavioural or design scientists use, we propose that this should be the case.

While we are here focussed on DSR as explicated and used within the field of Information Systems, Venable (2010) asserted that DSR applies in all business disciplines. Among others he notes that Van Aken (2004; 2005; 2007) advocates applying DSR within the field of Management.

This paper is organised as follows. The next section reviews the relevant literature on DSR. Section three applies a DSR lens to the goals and objectives of research methods, research methods as purposeful (designed) artefacts, requirements for evaluation of research methods, and design theories in research methods. Section four summarises the paper and identifies areas for further research.

2. Key aspects of design science research for research methodology

In this section we consider relevant literature from the field of Information Systems (IS) on Design Science Research (DSR).

We will define DSR as “Research that invents a new purposeful artefact to address a generalised type of problem and evaluates its utility for solving problems of that type”. There are four important concepts in this definition. First, by purposeful artefact, we mean any kind of artefact designed to achieve some human purpose. A purposeful artefact can be a product or a process; it can be a technology, a tool, a methodology, a technique, a procedure, a combination of any of these, or any other means for achieving some purpose. Second, by invention we mean creation, design, improvement, or adaptation, not just ‘pure’ invention from scratch (which may not exist anyway). Third, by addressing a generalised type of problem (rather than only a single specific, situated problem), the resulting artefact (and knowledge produced about it and its utility) can be applied over and over for various occurrences of that problem type. Without this generalised outcome, the scope of the resulting knowledge and the purposeful artefact would be too narrow to be useful in other contexts. Fourth, without evaluation, one has no evidence that the produced artefact has utility for solving problems of the type identified. Rigorous evaluation provides evidence that the artefact works and the knowledge created is true and useful. Without evaluation, there is no science in Design Science Research.

As an example of DSR, consider the development of the now commonplace balanced scorecard (Kaplan and Norton 1992; Kaplan and Norton 1996). While the creation of the balanced scorecard is in the DSR paradigm, Kaplan and Norton undoubtedly did not consider it to be.

The balanced scorecard approach was intended to provide a more holistic, comprehensive, yet still not overly complex picture of firm performance that incorporated not only financial, but also operational performance, to enable better firm management in the long run. The resulting artefact, the balanced scorecard, gives the manager “a fast, but comprehensive view of the business” (Kaplan and Norton 1992, p. 71).

The remainder of this section considers the nature of purposeful (design) artefacts, design theories as formal statements of the knowledge created by DSR, and evaluation in DSR.

2.1 Purposeful (design) artefacts

DSR produces one or more purposeful artefacts. March and Smith (1995) were the first in the IS field to establish the terminology of Design Science. They identified four kinds of what they call ‘design artefacts’ as the outputs of DSR: constructs, models, methods, and instantiations. Constructs are concepts that are used to describe a problem and specify its solution. Models are defined by March and Smith (1995) as “a set of propositions or statements expressing relationships among constructs” (p. 256) that can be used to “represent situations as problem and solution statements” (p. 256). They define methods as “a set of steps (an algorithm or guideline) used to perform a task” (p. 257). Finally,
they define instantiations as “the realization of an artifact in its environment” (p. 258). Interestingly, the above four artefacts do not include “design” per se. However, the model artefact could be used for this, in that it includes “solution statements” (p. 256). A design can then be conceived of as different constructs (the design components) and statements about relationships between them.

For example, consider the balanced scorecard (Kaplan and Norton 1992; Kaplan and Norton 1996) further. Considering March and Smith’s (1995) four kinds of artefacts, there are a number of key constructs or concepts, including the four proposed areas/perspectives on performance measures (financial, customer, internal business, and innovation and learning). Other constructs include the goals and measures to be identified within each perspective. The main model in the balanced scorecard is the view that integrates the four perspectives and shows the links between each of the four perspectives, as well as the goals and their measures within each perspective. Of course, another key concept is that it is up to the firm management to come up with the goals and measures that are appropriate for their organisation. In their original paper, Kaplan and Norton (1992) did not provide a method for how to apply the balanced scorecard, but it is implicit. They did describe an example part of an instantiation of their approach, as carried out at ECI.

Gregor and Jones (2007) distinguish two different kinds of purposeful artefacts that can be designed: product artefacts and process artefacts. A product artefact is a thing, such as a tool, an object, or a system. It can be either physical (e.g. a computer) or abstract (e.g. a diagram notation). A process artefact is similar to a method (March and Smith 1995). Of course, a purposeful artefact may be designed to include both process and product.

The balanced scorecard artefact contains both product and process aspects. The product is a scorecard itself and the constructs/concepts behind it. The process is the method by which the relevant measures of the four perspectives are decided and implemented.

Venable et al (2012) also distinguish between purely technical artefacts (which are admittedly fairly rare, such as a computer network protocol, which has no human users, and socio-technical artefacts, which have human users, possibly with complex social implications from their adoption and use. Importantly, socio-technical artefacts present a much richer milieu concerning the values and interests of the stakeholders as well as complications for the evaluation of the artefact (Venable, Pries-Heje et al. 2012).

The balanced scorecard is clearly a socio-technical artefact, since it has both individual usability aspects as well as group and social aspects in its use and the process of adapting it for a particular organisation.

2.2 Design theory

Design Theories are formalisations of the knowledge that results from Design Science Research. It is generally considered that all high quality research should generate, evaluate, or refine theory.

Walls et al (1992) argued that a proper design theory would have seven components as shown in figure 1: meta-requirements (a general set of requirements that a generalised solution would address), meta-design (a generalised design that could be adapted to a particular problematic situation), a design method (for adapting the meta-design to the particular problematic situation), kernel theories informing the meta-design (according to the meta requirements), kernel theories informing the design method (according to the meta-re quirements and meta-design), testable hypotheses to test the meta-design and testable hypotheses to test the design method. They also state that design theories are prescriptive in guiding designers as to what they should do.

Gregor and Jones (2007) developed an alternative, but similar form of design theory with eight parts: Purpose and scope, Constructs, Principles of form and function, Artefact mutability, Testable propositions, Justificatory knowledge, Principles of implementation, and an expository instantiation (optional).

Other authors have proposed simpler structures for design theories, focussing on meta-requirements and meta-design. Venable (2006) asserted that kernel theories and testable hypotheses are not part of a design theory per se and that a design method is not required to form a design theory. He advocated a simpler, non-prescriptive form of theory (which he termed a “utility theory”) containing
only (1) the problem space to which a design theory applies (similar to meta-requirements), (2) a solution space describing a design (similar to meta-design), and (3) a relationship between them asserting that the design has utility of some kind (efficiency, effectiveness, etc.) with respect to the problem space. Relationships between components of the problem space and between components of the solution space should also be part of the design/utility theory. Similarly, Baskerville and Pries-Heje (2010) proposed explanatory design theory, which includes three similar parts: (1) General Requirements (similar to meta-requirements), (2) General Components (similar to meta-design), and (3) a relationship between the two which means that the general components together satisfy the general requirements. In essence, the theory explains how to meet the general requirements, i.e. by applying the general components.

Figure 1: Structure of a design theory (Walls et al, 1992)

Since Kaplan and Norton (Kaplan and Norton 1992; Kaplan and Norton 1996) did not explicitly follow a DSR approach in their development of the balanced scorecard and the ideas of design theories had not yet been invented, it is not surprising that they did not explicitly propose a design theory. However, we can reconstruct at least part of a design theory post hoc, using the structure of Walls et al (1992).

Table 1 shows the different components of design theories according to the different works reviewed above in the left column and in the right column shows what each component of a design theory for the balanced scorecard might look like if Kaplan and Norton had developed one. Similar design theory components from the descriptions above are shown in the same row, The notations “WWES”, “G&J”, “V”, and “B&PH” identify the paper authors who developed the identified components.

Considering Table 1, the requirements shown in the first row are ones that need to be filled in many organisations; hence they are termed meta-requirements or generalised requirements. The particular goals and measures are left non-specific and have to be decided my managers for each particular organisation and their current situation, hence they are part of a meta-design or generalised design, rather than a specific design of a balanced scorecard for a particular organisation. The design method is left implicit in Kaplan and Norton (1992), but would likely require the steps shown. Kaplan and Norton (1992) somewhat notoriously did not provide references to any literature on prior art or to the requirements that they were trying to achieve. Therefore they did not provide any kernel theory. However, we can infer relevant literature which might (better) inform the meta-design and design method. Finally, we surmise some example testable hypotheses or propositions, but many others are possible.

Table 1: Hypothetical design theory for the balanced scorecard

<table>
<thead>
<tr>
<th>Design Theory Component (Source)</th>
<th>Hypothesised Design Theory for Balanced Scorecard (Kaplan and Norton 1992)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-requirements (WWES)</td>
<td>Enhance the ability of managers to manage a firm in the long run</td>
</tr>
<tr>
<td>Purpose and scope (G&amp;J)</td>
<td>Enhance the ability of managers to manage both financial and operational</td>
</tr>
<tr>
<td>Problem Space (V)</td>
<td>aspects of a firm</td>
</tr>
<tr>
<td>General requirements (B&amp;PH)</td>
<td>Reduce the complexity of managing a firm</td>
</tr>
<tr>
<td></td>
<td>Develop an integrated and holistic view of the firm</td>
</tr>
<tr>
<td></td>
<td>Support adapting the approach to specific organisations</td>
</tr>
<tr>
<td>Meta-design (WWES)</td>
<td>Four perspectives (financial, customer, internal business, and innovation and</td>
</tr>
<tr>
<td>Constructs, Principles of form</td>
<td>learning)</td>
</tr>
<tr>
<td>and function,</td>
<td>The links between the four perspectives (only the financial and innovation and</td>
</tr>
</tbody>
</table>
Design Theory Component (Source) | Hypothesised Design Theory for Balanced Scorecard (Kaplan and Norton 1992)
--- | ---
Artefact mutability, and Principles of Implementation (G&J) | learning perspectives are not linked to each other
Solution space (V) | Goals should be included within each perspective
General components (B&PH) | One or more measures for each goal
Guideline that there should be no more than 20 measures overall to ensure sufficient simplicity and ability to gain oversight of the entire firm
Design Method (WWES) | Decide the goals for each perspective
 | Decide the measures for each goal
 | Decide how to obtain the measures
 | Decide how to present the measures to managers
 | Build or buy and implement an activity system to provide the information
 | Use the measures to manage the firm
Meta-Design Kernel Theories (WWES) or Justificatory Knowledge (G&J) | Could be drawn from extant literature … on why managing is difficult
 | informing the design of the four different perspectives about extant goals and measures that have been identified and used in the different perspectives
Design Method Kernel Theories (WWES) or Justificatory Knowledge (G&J) | Could be drawn from extant literature on … how to collaborate on decision making about management
 | how to present complex information simply
Meta-Design Testable Product Hypotheses (WWES) and Propositions (G&J) | Possible testable hypotheses or propositions re. the meta-design might be …
 | “Use of the balanced scorecard reduces complexity in managing a firm.”
 | “Use of the balanced scorecard improves long term firm management.”
Design Method Testable Process Hypotheses (WWES) and Propositions (G&J) | Possible testable hypotheses or propositions re. the design method might be …
 | “The design method enables effective implementation of balanced scorecards.”
Instantiation (G&J) | A partial example instantiation was provided based on the ECI case study.

2.3 Evaluation

Evaluation is a key activity in Design Science Research (Nunamaker et al. 1991; March and Smith 1995; Hevner et al. 2004; Vaishnavi and Kuechler 2004; Venable 2006; 2006; Peffers et al. 2008). Evaluation is the activity that makes DSR a science. Without it, asserted design theories, utility theories, or technological rules are just untested theories, conjectures, and hypotheses; there is no evidence to back them up.

Venable et al (2012) identify five different purposes for evaluation in DSR:

- Evaluate an instantiation of a purposeful (designed) artefact to establish its utility (including efficacy, effectiveness, and efficiency – or lack thereof) for achieving its stated purpose
- Evaluate the formalized knowledge about a purposeful artefact’s utility for achieving its purpose
- Evaluate a purposeful artefact or formalized knowledge about it in comparison to other extant purposeful artefacts’ utility to achieve a similar purpose
- Evaluate a purposeful artefact or formalized knowledge about it for side effects or undesirable consequences of its use
- Evaluate a purposeful artefact formatively to identify weaknesses and areas of improvement while still under development

There are four questions we should answer in designing an evaluation:

- What is the evaluand (the artefact being evaluated)?
- What aspect or quality of the evaluand is to be evaluated? Artefacts are evaluated for their fitness to purpose. Purposes are extremely varied, but one can identify some key characteristics, such as Checkland’s five E’s : efficiency, effectiveness, efficacy, (a)esthetics, and ethicality (Checkland and Scholes 1990), which are aspects of utility.
What kind and how much rigour are required of the evaluation? Venable et al (2012) identify two senses of rigour. The first sense is whether an improvement or achievement of purpose can be attributed to the evaluand (i.e. its efficacy), or alternatively whether it could be caused by some confounding variable. The second sense is whether the evaluation can work within the myriad complicating aspects of real use, i.e. its effectiveness.

What evaluation method(s) should be used? Hevner et al (2004) suggest analytical, case study, experimental, field study, and simulation. Venable (2006) further suggests action research and distinguishes between artificial evaluation and naturalistic evaluation. But which evaluation method(s) should be used?

Venable et al (2012) provide extensive advice on evaluation method choice and design based on a framework matching the prioritised goals of the evaluation to strategies and then to specific relevant methods. Strategies consider whether evaluation should be ex ante (before instantiation) or ex post (after instantiation) and whether it should be naturalistic (real users using the real purposeful artefact on a real task (Sun and Kantor 2006)) or artificial (lacking one of the three realities of Sun and Kantor (2006)) or a combination of these.

Considering evaluation of the balanced scorecard, the evaluand (purposeful artefact evaluated) could be the balanced scorecard itself, some more focussed component part of it, such as different ways to represent the balanced scorecard, and/or the method for designing and using a balanced scorecard in a particular organisation.

Kaplan and Norton (1992) provided a fairly minimal evaluation of the overall approach by applying the method in several different case study organisations. Because this was uncontrolled, it provided only subjective, anecdotal evidence of the effectiveness of the purposeful artefact, not of its efficacy. Anecdotal evidence was also provided of the efficiency of the method. The method was also not compared to other methods.

Other researchers have also evaluated the balanced scorecard approach, e.g. Mooraj et al (1999) and Malina and Selto (2001). These evaluations were more rigorously performed, but still used case studies and largely considered subjective opinions (e.g. the perceived attributes of the balanced scorecard for strategy formulation and communication (Mooraj et al. 1999)). Both evaluations were generally positive, but could not be said to rigorously evaluate the efficacy of the balanced scorecard.

Unfortunately, constructing a fully-controlled evaluation of the efficacy of the balanced scorecard is probably impossible; the key anticipated benefit (dependent variable) – improved firm performance in the long term – is subject to many confounding variables.

Other desired benefits, such as improved ability to manage, improved ability to deal with both financial and operational performance in an integrated and holistic way, and reduced complexity, are all subjective in nature and could be evaluated by asking people about their subjective opinions on whether, how well, or how much better the purposes are achieved.

3. Analysis of research methods using a design science research lens

This section considers what lessons may be learned from applying Design Science Research to research methods and how DSR might improve the rigour of research methods. We look at four main areas: goals (and constraints) of research methods, design artefacts in research methods, the need for evaluation of research methods, and design theory in research methods.

As an example for the purpose of this paper, we will use the Partial Least Squares (PLS) method (Wold 1974; Wold 1982; Lohmöller 1989; Chin 1998; Chin and Newsted 1999) for analysing quantitative data in the context of Structured Equation Modelling (SEM or PLS-SEM). One reason for this is that the use of PLS for business research is a matter of extensive debate (Marcoulides and Saunders 2006; Marcoulides et al. 2009; Gefen et al. 2011; Hair et al. 2011; Ringle et al. 2012; Rönkkö et al. 2012).
3.1 Goals and purposes of research methods

This section identifies and discusses some of the various goals to be met by research methods. Goals include both the primary goals as well as practical goals that enable their effective use. We discuss those first, then implications from a DSR perspective.

3.1.1 Primary goals

The primary goals of research methods include both general goals and goals specific to the research project. Rigour and relevance are general goals desired of all research and research methods. Research methods must also be appropriate for answering the specific research question or achieving the specific goal of the research.

Rigour is concerned with the reliability of research. Properly following a research method should ensure that the research findings are correct or that the probability of the findings being incorrect is sufficiently low.

Relevance is concerned with whether the findings are useful – either to theory and further research or to practice. While following a research method may not ensure relevance, some research methods may be more amenable to investigating topics that are relevant.

Research questions may be of several kinds or types. Some research questions are concerned with theory testing (and extension). Some are concerned with theory building. Some are concerned with understanding and explanation while others are concerned with prediction (cf. Gregor 2006). Some are concerned with evaluation. Some are concerned with inventing new technologies. Some are concerned with human emancipation. Different research methods are more or less appropriate for answering different kinds of research questions.

A primary goal of PLS-SEM is to support rigorous analysis of quantitative data. The relevance of its results are mostly dependent on the relevance of the questions and the data collected, but, as a positivist, reductionist approach dealing with quantitative data, there are many areas where more interpretive approaches may bear more fruit. In general, it is recommended that PLS-SEM be applied for exploratory research (Gefen et al. 2011) and theory building and prediction rather than theory testing. PLS is most relevant and appropriate for social and behavioural research, including business research (Gefen et al. 2011; Hair et al. 2011).

PLS-SEM (along with other SEM forms, i.e. co-variance-based SEM or CB-SEM) also has a key advantage in its support for latent variables (or concepts), which cannot be directly measured.

3.1.2 Practical goals

There are many practical goals for research methods that affect their ability to be applied and to achieve the primary research goals given above. Among them are ease of learning, ease of use, cost, time required, facilities or other resources required, access to research subjects, ability to deal with more people and research subjects, and depth of engagement with research subjects. These goals are often in the form of constraints – constraints on available researcher time, expertise, funding, access to research subjects, and others. The list is probably not comprehensive.

A key practical advantage for PLS is its (supposed) ability to work with small(er) sample sizes. However, this is under much debate and widely criticised (Marcoulides and Saunders 2006; Marcoulides et al. 2009; Gefen et al. 2011; Ringle et al. 2012; Rönkkö et al. 2012). Smaller sample size requirements is often a reason for using PLS, with over 1/3 of papers using PLS published in MIS Quarterly citing this as a reason for choosing PLS (Ringle et al. 2012).

PLS, particularly with the software support available today (e.g. SmartPLS), is quite easy to learn and use, so in some sense is suited to novice researchers with limited time resources. However, it can also be argued that the constraints and limitations of the technique, as well as reporting standards (Gefen et al. 2011; Ringle et al. 2012), are quite difficult to learn as we shall see below.
3.1.3 Lessons from DSR

The goals above are analogous to the DSR idea of meta-requirements. The different research methods are analogous to the DSR idea of meta-designs. The different goals/meta-requirements form contingencies that are addressed with differing levels of utility by the different research methods.

A key issue is that the number of goals and constraints is fairly large and to some extent they conflict. Evaluating the different contingencies is complex and difficult as it is a combinatorial problem without hard yes/no answers. However, the provision of a clear system of design theories with research goals as the contingencies could aid in the selection of an appropriate research method for a particular research situation. See for example Venable et al (2012) for a framework and method to aid researchers in the selection and design of evaluation methods for DSR purposeful artefacts.

Moreover, there may be room to develop a system or tool to aid in the selection of research methods based on the specific goals and priorities of a research project. For example, the approach described in Baskerville and Pries-Heje (2008) for helping managers to select an appropriate change management methodology might also be applied to a system to help selection of research methods. Given that research questions sometimes evolve based on the feasibility of researching them, such a tool might also be modified to support choosing/designing research questions.

3.2 Design artefacts in research methods

As described in section 2.1, design artefacts are distinguished as product and process artefacts (Gregor and Jones 2007), as well as classified as constructs, models (including designs), methods, and instantiations (March and Smith 1995; Hevner et al. 2004).

What kind of design artefact is a research method? A research method is clearly a process or method artefact. Research methods often (always?) make use of various tools, techniques, and frameworks for organising and doing the research work. These are clearly product artefacts. Descriptions of how to use a tool, technique, or framework are in turn process artefacts.

Importantly, tools and techniques may sometimes be used in other methods than where they were developed. They have an existence of their own and could be used in other ways. They have their own goals and instructions for how to use and adapt them. Any attempt to conduct DSR on research methods needs to consider the possibility that research tools and techniques have uses both as part of and outside of any particular research method.

Considering PLS, there are many artefacts which could be evaluated. PLS is an algorithm (process), implemented in software (product), with a user interface (product), and also procedures (process) and standards (product) for its application and submission of results for publication. One could evaluate the PLS algorithms, the correctness and usability of PLS software, how easy it is to learn and apply the standards or practice, whether PLS achieves the rigour that is required when applied with its constraints and standards (its efficacy), the efficiency of its application by people (time, resources used, etc.), and the effectiveness of its use (obtaining rigour in actual practice by researchers).

3.3 Evaluation of research methods

In the introduction, we asserted that research methods have evolved in a non-rigorous way, without the benefit of the ideas of Design Science Research. A key aspect of DSR is evaluation of new technologies.

3.3.1 Why evaluate research methods?

Research methods are tremendously important to the progress of human knowledge and the improvement of the human condition. Currently, we lack evidence of the effectiveness, efficiency, and efficacy of research methods. Some evidence clearly is provided by the evident progress of human knowledge, but enabled by which research methods? Other evidence has been developed through rational argument in research essays. However, without more rigorous evidence, how do we really know that research methods are (sufficiently) effective, efficacious, or efficient?

We also lack evidence about whether research methods achieve other desirable, practical goals – e.g. ease of learning, ease of use, and time and resource requirements.
Moreover, we especially lack evidence about the relative efficiency and effectiveness of research methods, i.e. compared to each other.

In order to make better judgments in the choice of research methods, it would be very helpful to have better evidence than our currently ad hoc assertions about them.

For example, Rönkkö et al (2012) recently evaluated the use of PLS for theory testing who found that many models developed using PLS and published in top IS journals may in fact be incorrect. Of course the fault may lie in the application of PLS rather than the PLS algorithms or their software implementations. Importantly, their research shows that, whether the PLS technique would work (determine correct models) if applied correctly, the PLS technique as recently applied in the IS field, is (or has been) ineffective in practice, as used and published by IS researchers, reviewers, and journal editors.

3.3.2 How could we evaluate research methods?

Like other designed artefacts, research methods should be evaluated for their utility in achieving their goals, including primary goals of rigour and relevance, suitability to type of research questions or research domain and topic, as well as secondary, practical goals.

Rigour as it is assessed in the reviewing process is a rather subjective and often assessed by simply confirming that an established research method has been followed, assuming that the outcome will therefore be correct. Sometimes, sets of criteria are set out for what makes research rigorous, trying to make the criteria as specific and clear as possible, even objective if possible. But what if the method itself or the criteria are wrong?

The most convincing evidence of the rigour of a research method would come from the detection of errors in findings and determining error rates. Where errors are found in findings, the research method itself is in some way responsible (after all, it didn’t prevent the error), but additional causes are possible (e.g. misapplication of the method). Finding errors would require triangulation of findings through other research, which may not be possible or thought to be worth the research expenditure.

For some research methods, especially quantitative ones, the efficacy of the method to provide rigour can be demonstrated through mathematical analysis or proof.

Alternatively, research conducted could be subjected to critical analysis and ratings of rigour. The rigour ratings could then be compared and correlated across projects using different research methods. Unfortunately, mixed- and multi-method approaches would complicate these analyses. It may also be helpful analyse failed research projects, where outcomes are not accepted for publication. One might also consider whether a research method correlates with high or low quality journals.

To assess the performance/utility of research methods in achieving goals in other, practical areas one could survey users of various research methods about their perceptions of ease of learning and ease of use of research methods, tools, and techniques. One could evaluate resource consumption by gathering data on time required, resources consumed, costs, etc. and correlating them with research methods.

The above suggestions relate to extant research methods, but what about the development of new methods where there isn’t a history of application? Where a method is new, evaluation is largely limited to analytical kinds of evaluation, e.g. rational argument, surveys or focus groups with experts in research methodology, or possibly some form of simulation. Empirical, naturalistic evaluation will likely be very limited at first, with limited trials, probably using qualitative evaluation methods.

Considering evaluation of PLS, Rönkkö et al (2012) did not evaluate the process by which PLS was applied; instead it examined the results of the application of PLS and publication of PLS results. It did so be re-running the analyses, but with modified data, including (1) switching the data between two variables and (2) substituting completely random data for two variables. They found that in a significant portion of the models re-analysed with incorrect data, the model was still supported by the analysis. Since models with incorrect data were supported, presumably the models themselves may be incorrect since the evidence through the use of PLS is extremely weak.
3.4 Design theory for research methods

Design theories would be important in a Design Science of Research Methods.

3.4.1 Why design theories of research methods?

Venable (2006) observed that theory is the way that academics communicate with each other about research. Clear, formalised statements of theory are needed to reduce misinterpretation and to allow other researchers to accurately research the same topics. They are also important when developing new research methods as one needs to be able to clearly articulate how a new research method is different from existing ones.

Clear statements of design theories may also be useful for practitioners – in this case the practitioners are researchers.

For these reasons, we propose that formal design theories could and should be developed for every research method. Doing so would aid the on-going development and progress in research methods and in the long run should increase our ability to generate new, valid knowledge.

3.4.2 What should be in a design theory for a research method?

At a minimum, a design theory for a research method needs three things: (1) meta-requirements (Walls et al. 1992), a model of the problem space (Venable 2006) to be addressed by the research method, a purpose and scope (Gregor and Jones 2007), or generalised requirements (Baskerville and Pries-Heje 2010), (2) a meta-design (Walls et al. 1992), a model of the solution space (Venable 2006), principles of form and function (Gregor and Jones 2007), or a generalised design (Baskerville and Pries-Heje 2010), and (3) a prescriptive statement (Walls et al. 1992; Gregor and Jones 2007), an asserted utility relationship (Venable 2006), or an explanatory relationship (Baskerville and Pries-Heje 2010) between the first two. Other aspects of design theories are optional.

All aspects of design theories need to be stated as precise, well-defined constructs. The meta-requirements/model of the problem space/purpose and scope should include and clearly state the goals (and possibly the constraints) for which the research method is suited (or possibly the level to which they are suited, such as low cost). The primary goals that should be addressed include research rigour and relevance and the level of each that can be achieved or expected, the type(s) of research question for which the method is suited, and the domain(s) for which the research method is suited. Optional goals may also be addressed (we recommend this) by statements about the anticipated levels of ease of learning, ease of use, cost, time required, facilities or other resources required, access to research subjects, ability to deal with more people and research subjects, and depth of engagement with research subjects.

The meta-design/model of the solution space/principles of form and function should include a detailed description of the research method’s process, including steps, actions to take, tasks, decisions, and iterations, as well as tools and techniques to use, and when and how to use them. Descriptions of tools and techniques may refer to other design theories for detail about them.

Table 2 gives an example of a design theory for the PLS approach to quantitative data analysis. In this case, we take on the results of the Rönkkö et al. (2012) evaluation and propose some improvements to PLS as it is currently implemented.

Table 2: Hypothetical (partial) design theory for the partial least squares method

<table>
<thead>
<tr>
<th>Design Theory Component (Source)</th>
<th>Hypothesised Design Theory (partial) for an improved Partial Least Squares method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-requirements (WWES)</td>
<td>Support use of latent variables and both formative and reflective measures.</td>
</tr>
<tr>
<td>Purpose and scope (G&amp;J)</td>
<td>Support use of small data sets (within minimum constraints).</td>
</tr>
<tr>
<td>Problem Space (V)</td>
<td>Support easy building and testing of models (efficiency and ease of use).</td>
</tr>
<tr>
<td>General requirements (B&amp;PH)</td>
<td>Support easy calculation and application of constraints to ensure adequate rigour</td>
</tr>
<tr>
<td></td>
<td>(i.e. correctness) of results.</td>
</tr>
</tbody>
</table>
4. Conclusion

Research Methods are designed artefacts. Currently, there is no adequate basis for assuring the utility and rigour of research methods. The evolution and development of methods needs a more rational basis. This paper proposes that DSR could (and should) provide such a basis. If we are to be certain that research methods are reliable and achieve their purpose, they should be developed with a rigorous DSR approach. The rigour would come from (a) more precise statements of design theories of research methods and (b) more rigorous evaluation of research methods.

In order to progress toward a rigorous DSR approach to research methods, we have applied DSR concepts to research methods, tools, and techniques. We have identified the primary and secondary (practical) goals of research methods, which would be part of design theories about research methods and serve as criteria for evaluation of the utility of research methods. We have also used DSR concepts to identify types of design artefacts in research methods, which would define the research method in a design theory. The configuration of constructs and artefacts that make up a research method are what are evaluated for how well they fulfil the research method’s goals. This understanding should also lead to more consistent evaluation of research methods.

Applying a DSR approach to research about research methods should enhance the quality of research methods and therefore the quality of the research conducted using them.

What we have proposed in this paper is only a first step. Work is now needed to analyse existing research methods to develop clear design theories about them. The basis for this already exists in the literature; it simply needs critical reading and formalising in design theories. Consideration also is
needed re-designing and carrying out appropriate evaluations of research methods, to provide evidence of their effectiveness, efficiency, efficacy, etc. in achieving their disparate purposes. Once a body of literature on design theories and evaluations of research methods is developed, there would be a clearer basis for further research on research methods.

References


