Issues in Social and Environmental Accounting
Vol. 5, No. 1 December 2011
Pp 25-64

Quality in Qualitative Studies: The Case of Validity, Reliability and Generalizability*)

Azham Md. Ali[1**)]
Hamidah Yusof
Faculty of Management and Economics
Universiti Pendidikan Sultan Idris
Malaysia

Abstract

The concern over the subject of quality judgement of qualitative research has led to the emergence of several varied positions. Whichever labels used in describing the positions, each has a distinctive character. To attempt to an understanding of the meaning of the quality criteria of validity, reliability and generalizability in qualitative research, this paper provides a review of relevant literature of recent years. To check on the extent of their reporting in research papers, a total of 15 accounting research recently published in top-ranked accounting journals have been selected. Except for one paper, the rest has failed to provide detailed discussion on the validity, reliability and generalizability of their research. But this would not necessarily mean that the concerned researchers have disregarded the aspects of validity, reliability and generalizability during the conduct of their research. Perhaps the only way to find out the truth is by asking the researchers themselves the relevant questions.

Keywords: *qualitative research, quality, validity, reliability, generalizability, accounting*

[1] Azham Md. Ali is currently a Professor of Accounting and Finance at the Faculty of Management and Economics at the Universiti Pendidikan Sultan Idris, Tanjung Malim Perak. Prior to his current appointment, he was attached to the Universiti Utara Malaysia, Sintok Kedah for which he had served for close to 25 years. His areas of interest are auditing, international accounting, accounting history and qualitative research. To date, he has published in numerous local and international journals and presented papers in various conferences and seminars.

*) A slightly different version of the paper with the same title was presented as Invited Paper II for **The Second National Seminar on Qualitative Research**, 4-5 December 2006 at Kompleks Dewan Kuliah Sains, Fakulti Sains, Universiti Malaya. The seminar's organizers are the Qualitative Research Association of Malaysia (QRAM) and Institut Pengurusan Penyelidikan dan Perundingan, Universiti Malaya.

**) The following people from the Faculty of Accountancy, Universiti Utara Malaysia (UUM) have been quite supportive in different ways over the successful completion of the original version of the paper (which was presented as Invited Paper II for The Second National Seminar on Qualitative Research, 4-5 December 2006 organized by the Qualitative Research Association of Malaysia and the Institute of Research Management and Monitoring, University of Malaya): Noor Yusni Md. Yusuf, Azharudin Ali, Dr. Nor Aziah Abd. Manaf and Dr. Mohamad Hisyam Selamat. Thus, the main author is very much appreciative of their support. Since the opportunity to work on the paper has come from Dr. Siti Nabiha Abdul Khalid from Universiti Sains Malaysia (USM) and Dr. Aini Hassan from Universiti Malaya (UM), the main author is also very much grateful to them. Finally, a lot of gratitude is for the author's former PhD supervisor Prof. Stuart Turley from the University of Manchester, UK who in 1995 had been quite understanding enough in allowing him to change the research approach (following the change in the research question raised) from quantitative to qualitative and which has led him to come to know the very interesting world of qualitative inquiry.

> At present, there is an observable trend that qualitative research attracts students as well as practitioners who are not at ease with demanding quantitative procedures and who expect that in qualitative research they can substitute methodological sophistication with common-sense. The more fashionable qualitative research has become in different fields, like social work, business administration, marketing and consulting, the more we face a quality problem. Any investigation which does not make use of statistical procedures is called "qualitative" nowadays, as if this were a quality label in itself. As many have already called for: we need to define criteria by which we can distinguish "good" from "bad" qualitative research, be it "validity" and "reliability" or other concepts like "consistency" or "adequacy of meaning" or even others.
>
> Eberle (2005, p. 4)

1. Introduction

Quality concerns play a central role throughout all steps of a research process from the inception of a research question and data collection, to the analysis and presentation of research findings. Following this line of thought, there is no question that quality is an integral and explicit part in qualitative inquiry. However, the concern over the subject of quality judgement of qualitative research has led to the emergence of several varied positions. See Table 1 below.

Table 1
Quality Judgement Positions

Literature	Quality Position
Denzin (1995)	positivism, postpositivism, postmodern and postfoundational
Onwuegbuzie (2002)	positivists, postpositivists, poststructuralists and postmodernists
Lazaraton (2003)	foundationlism (positivism/rationalism), quasi-foundationalism (postpositivism, constructivism) and nonfoundationalism (postmodernism)

In sentence form, these quality positions may be understood as the followings:

- There is *only one way* to judge the quality of qualitative studies which is the same for any type of scientific inquiry: the criteria of reliability, internal and external validity and objectivity.
- There is *no way* to judge the quality of qualitative studies.
- The way to judge the quality of qualitative studies is through *modifying* quantitative research criteria to produce criteria fit for qualitative research context.
- The way to judge the quality of qualitative studies is by developing a *unique* set of criteria.

Other than these four quality positions, a qualitative study may also be judged on its quality through

- the conduct of a deductive quantitative study to test out the emerging theory.

This paper is an attempt to understand the meaning of the quality criteria of validity, reliability and generalizability in qualitative research. Following that attempt, the paper focuses on the reporting of these and other quality criteria in published studies. In order to come to some understanding (or perhaps utter confusion?) of what these three terms stand for, there is the review of selected recent literature. The review is done with no intention whatsoever to provide anything new or different in regard to the meaning of validity, reliability and generalizability of qualitative studies as compared to those offered by various parties to date. To name just four of these parties, these would be Davies and Dodd (2002), Golafshani (2003), Kvale (1995), and Whittermore, Chase and Mandle (2001).

Hence, unlike Davies and Dodd (2002), Golafshani (2003) and Kvale (1995), the paper does not take a critical focus on the inadequacy of applying a quantitative concept of rigour to evaluate qualitative research. In short, it does not attempt to redefine the concepts of reliability, validity and triangulation. Also, the paper does not see the need to add to or subtract from the present plethora of meanings of validity, reliability and generalizability of qualitative studies. This is done so that the paper is *never in danger of introducing yet more confusion into the already turbulent waters of the meaning of quality in qualitative studies! Related to this, note the following coming from Creswell and Miller (2000, p. 124):*

> … readers are treated to a confusing array of terms for validity, including authenticity, goodness, verisimilitude, adequacy, trustworthiness, plausibility, validity, validation, and credibility. Various authors have constructed diverse typologies of validity (e.g., Maxwell's five types, 1992; Lather's four frames, 1993; and Schwandt's four positions, 1997). It is little wonder that Donmoyer (1996), who wrote and editorial on validity in the *Educational Researcher, commented on the diverse perspectives of validity… Novice researchers, in particular, can become increasingly perplexed in attempting to understand the notion of validity in qualitative inquiry.*

So, while the paper unlike so many others such as Whittemore et al. (2001) does not provide new or different meanings to what is currently understood to be validity, reliability and generalizability in qualitative inquiry, it is on the other hand pretty much like Morse, Barrett, Mayan, Olson and Spiers (2002) which propounds the implementation of the so called verification strategies during the conduct of a qualitative inquiry to ensure the attainment of rigor or quality in such work. That said, the paper does not make the same kind of stance as that of Morse et al. (2002): a return to terminology for ensuring rigor in qualitative inquiry that is used by mainstream science, i.e., validity and reliability. In the final analysis, the ultimate aim of the paper is as mentioned succinctly by Seale (1999, p. 465):

> A lot of effort has been expended by methodologists over the years, trying to give some guidance to qualitative researchers in improving or judging the quality of qualitative research. You could say that all methodological writing is ultimately directed at such a goal, because the

> idea of writing about how one can do research is presumably aimed at giving other people some good ideas on how they might proceed with their own studies.

Apart from reviewing selected works on the subject of quality criteria of validity, reliability and generalizability which provides the picture that there is much confusion of what these terms actually mean in qualitative inquiry, the paper gives focus to the reporting of these quality criteria in a number of published research papers. One set of papers is comprised of qualitative accounting research recently published in top-ranked accounting journals. It appears that the paper provides the only attempt to date over such analysis of qualitative studies in accounting.

Perhaps, this is not surprising since over the last two decades there appears to be mere handful of published papers around which give focus on the subject of quality of qualitative studies in accounting. These would include Ahrens and Chapman (2006), Atkinson and Shaffir (1998), Lillis (2006), McKinnon (1988) and Modell (2005). In fact, in what appears to be one of the earliest writings on the subject matter of validity and reliability in qualitative accounting studies, the late McKinnon mentioned over two decades ago the following (McKinnon, 1988, p. 34):

> … field studies are frequently subjected to common and global criticisms of their apparent inability to attend to such research criteria as validity and reliability … many published field studies in accounting do not report how issues of validity and reliability are addressed.

Later, and in a more recent time, Irvine and Gaffikin (2006) had raised this matter by having it placed in a bigger context. They said (Irvine and Gaffikin, 2006, p. 115): "While much qualitative research has been undertaken within the discipline of accounting, little or no intention has been paid to the way in which that research has been conducted."

Aside from the analysis done on a selection of qualitative accounting papers, a total of twelve qualitative studies that are concerned with various other fields in social sciences have been analyzed in terms of their reporting of the quality criteria. Out of these twelve, five are PhD thesis while an additional three are journal writings whose discussion on the variety of quality criteria are concerned specifically with the use of specific research methods such as convergent interviews for data collection or analysis. These published works are chosen for analysis since they would provide some of the best portrayal of such reporting. Hence, they can be the models for those researchers who are looking for the acceptable ways in reporting the quality criteria which may or may not be the validity, reliability and generalizability that are the focus of the paper.

By having this analysis done together with that for a selected few accounting papers, it is hoped that the paper shall be able to contribute to improved practices in qualitative studies for those new in the field including doctoral candidates. For the doctoral students in particular they need no further reminder that demonstrating the trustworthiness of one's thesis is a requirement.

By and large, by focusing on the report-

ing of quality criteria in a total of 27 published studies so that budding researchers in qualitative inquiry may be assisted in their work, the paper attempts to be on the side propounded in Seale (1999) as opposed to those mentioned earlier as the varied positions on quality in qualitative studies. Specifically, Seale (1999) emphasizes the need for researchers to gain the so called "apprenticeship experiences" as opposed to "intense methodological awareness". He wrote (Seale, 1999, pp. 475-476):

> Methodological writing is of limited use to practicing social researchers, who are pursuing a craft occupation, in large part learned "on the job," through apprenticeship, experience, trial, and error rather than by studying general accounts of method … Intense methodological awareness, if engaged in too seriously, can create anxieties that hinder practice … people learn how to do research through apprenticeship experiences, fortunately possible to have by reading others' work rather than actually going and sitting at their feet (although this also can be useful). Any contemplation of other people's research work, if it involves thinking seriously about its strengths and weaknesses, can be this kind of vicarious apprenticeship experience.

All in all, Seale disagrees with the idea that philosophical, political, or theoretical positions ought to determine the decisions that social researchers make "on the ground" so that quality is underwritten by adherence to a particular position. Instead, in his view research practice should be conceived as relatively autonomous from such abstract and general considerations. In short, as far as he is concerned, particular craft skills such as member checking, accounting for negative instances, analytic induction, the uses of numbers, using low inference descriptors, the grounding of theory, deconstructive approaches, reflexive accounting and new textual forms of reporting do not have to be linked inextricably to particular philosophical or paradigm positions. In the final analysis, he considers a major threat to quality is the idea that research must be carried out under the burden of fulfilling some philosophical or methodological scheme. Instead, in his view, what should be the case is simply this (Seale, 1999, p. 472): "Practicing social researchers can learn to do good work from a variety of examples, done within different "moments," without needing to resolve methodological disputes before beginning their work."

Finally, it perhaps needs to be noted that the analysis done on the reporting of validity, reliability and generalizability in qualitative accounting papers is concerned with only a few of the relatively large number of these papers. A different sample might have given a somewhat different picture. Also, it perhaps needs to be stated out that the rudimentary analysis done for this paper on the reporting of quality criteria in fifteen qualitative studies from the accounting field and twelve others from various other fields in the arena of social sciences could very well move to the next level with the focus on the specific research strategies undertaken. Such further study with perhaps a much larger sample of published studies could follow in the footsteps of Barusch, Gringeri and George (2011) for the field of social work and Suter (2010) for personal relationship.

In Barusch et al. (2011), they use Creswell's (2007) eight strategies as the benchmark for rigour in assessing a random sample of 100 qualitative social work articles drawn from selected social work journals. As for Suter (2010), the strategies for validity that the so called "validity processes" are judged against come from various methodological writings in the field. A further study in the manner of Barusch et al. (2011) or that of Suter (2010) in a field such as accounting should be able to deepen understanding on the extent of rigour or quality in qualitative studies reported in published works.

The rest of the paper is divided into three sections. The next section covers the varied understanding of validity, reliability and generalizability in qualitative research – together and separately. The section which comes right after focuses on specific references made to these quality criteria in recent published papers in the accounting field. The last section is the discussion and conclusions. As part of this third and last section, there is the inclusion of the analysis done on the reporting of quality criteria by a total of twelve published works in social sciences.

2. Validity, Reliability and Generalizability

As mentioned in the introduction, there are varied positions over quality consideration for qualitative research. For many positivists, they feel that if a research does not satisfy several criteria, then it is not true research. These criteria are (Guba and Lincoln, 1994): internal validity, the degree to which the results can be attributed to treatment; external validity, the generalizability of the results; reliability, the extent to which the findings can be replicated; objectivity, the extent to which the findings are free from bias. As perhaps to be expected, those who do not consider themselves positivists oppose such views.

Janesick (1994, p. 217) challenges the notion that the "trinity of validity, generalisability and reliability", terms usually synonymous with the quantitative paradigm, should be strictly applied to all research. A decade later, Morgan and Drury (2003) who agree with her explain why: in qualitative research, the researcher is more interested in questioning and understanding the meaning and interpretation of phenomena. But the issue involving these quality criteria cannot easily be pushed away with such remark. This can be seen in the efforts of others working in qualitative inquiry in coming out with new terms to replace validity, reliability and other terms used in quantitative inquiry. Among the notable ones are Guba and Lincoln (1981), Lincoln and Guba (1985), and Whittemore et al. (2001).

In the case of Guba and Lincoln (1981), they propose that the criteria to reach the goal of trustworthiness in qualitative inquiry are credibility, fittingness, auditability and confirmability. These are as opposed to the criteria internal validity, external validity, reliability and objectivity to reach the goal of rigour in quantitative inquiry. A few years later, they suggest the criteria to now be credibility, transferability, dependability and confirmability (Lincoln and Guba, 1985). Not every one agrees however with these ideas of theirs. See for example Hammersley (1992), Kuzel and Engel (2001) and Yin (1994). Perhaps in

regard to Yin (1994) the opposition is not surprising. This is because he describes trustworthiness as a criterion to test the quality of research design and *not* a goal of the research as proposed by Guba and Lincoln (1981) and Lincoln and Guba (1985).

Also, over two decades later, there is a clear opposition among some parties in the use of these terms to describe the quality criteria in qualitative inquiry. Note the following coming from Morse et al. (2002, p. 3): "… the broad and abstract concepts of validity and reliability can be applied to all research because the goal of finding plausible and credible outcome explanations is central to all research." Later, in the same work, the following is mentioned (Morse et al., 2002, p. 14): "Our argument is based on the premise that the concepts of reliability and validity as overarching constructs can be appropriately used in all scientific paradigms because, as Kvale (1989) states, to validate is to investigate, to check, to question, and to theorise. All of these activities are integral components of qualitative inquiry that ensure rigor."

Also note the following which is mentioned more recently by Bergman and Coxon (2005, p. 3): "Quality considerations in empirical research tend to be addressed by the concepts "validity" and "reliability"… We will critically examine some possibilities of these concepts …" Next, in the attached footnote number 3, the following is what they say:

> It has often been suggested that these terms are inappropriate since they have emerged from a positivistic tradition. However, we argue that concerns about data quality transcend positivism; while we have nothing against coining new terms, particularly if this would avoid the conceptual baggage that may be attached to a certain terminology, we believe that we may want to examine existing tools before adding new terms to potentially similar concepts.

Later at the end of their writing, they mention the following (Bergman and Coxon, 2005, p. 13):

> Whether or not researchers coin their own terminology because they reject constructs that may have emerged from another epistemological tradition, or whether they begin their quality considerations by adopting the existing terminology is not important at this point. Instead, *it is the accountability of research practices through explicit description of research steps*, which allow an audience to judge the plausibility of a particular study and its findings. (Emphasis added.)

Validity

To understand what validity is in a research inquiry, one only needs to refer to the experts. But it seems that is not such a good idea as far as qualitative studies are concerned! For at least two main reasons. First, the experts themselves have failed to be consistent. Second, validity in qualitative inquiry comes about in so many different ways of understanding. When it concerns the first reason, two fine examples are Harry Wolcott and David Silverman. As for the second reason, the recent works by Spencer, Ritchie, Lewis and Dillon (2003) and Onwuegbuzie (2002) are

proof enough.

The first of two examples of experts in qualitative field who fail to be helpful over the subject of validity is Harry Wolcott who for over three decades, as the master ethnographer, has chartered an interpretive, postpositivist approach to the anthropology of educational practices (Denzin, 1995, pp. 181-182). But with his 1994 publication, calling for methods and texts that produce understanding, not validity, he states (Wolcott, 1994, p. 369): "I do not accept validity as a valid criterion for guiding or judging my work." In his 1990 publication earlier (as found in Onwuegbuzie, 2002, p. 9), he in fact raises the question as to whether validity is appropriate, legitimate or useful in qualitative studies. In his view, it seems validity interferes with his goal of understanding the underlying phenomenon. This has led Onwuegbuzie (2002, p. 9), in commenting over this stance of Wolcott, to mention the following: "According to Wolcott, understanding is a more fundamental concept for qualitative research than is validity. Consequently, he attempts to understand what is occurring rather than to convince his audience."

Nonetheless, Wolcott seems unable to dismiss validity outright. As pointed out by Spencer et al. (2003, p. 59), Wolcott is "[p]erhaps the most frequently cited example of someone who *apparently* rejects validity while retaining its underlying concerns …" (Emphasis is in the original.) Next, they specify that Wolcott (after saying that he cannot see any place for validity in his work) has in fact made suggestions on how to produce valid qualitative works. These are (Spencer et al., 2003, p. 59): listening more than talking; recording accurately; beginning writing early and sharing ideas with others in the setting; letting readers 'see for themselves'; reporting fully; being candid; seeking feedback; trying to achieve a balance through rigorous subjectivity; and, writing accurately.

In the case of Silverman, as compared to Wolcott, he provides a more recent example of resistance and confusion over the issue of validity in qualitative research. This is as far as Silverman (2001) is concerned that is comprised of among others a chapter on validity, reliability and generalization. As Kalekin-Fishman in her review of this work puts it (Kalekin-Fishman, 2001, p. 3):

> After acknowledging and detailing the weaknesses of the conceptualization of reliability and validity on the home ground of quantitative research, Silverman insists on applying these same terms for deciding the quality of every kind of research. He is, for example, adamant about testing for validity, i.e., for the "truthfulness" of evidence, even though he agrees with most qualitative theorists that this is highly dubious goal.

Indeed, if a check is made on the suggestions he made in the book regarding the attainment of validity in qualitative studies, one can find the followings: a reliance on theoretical models to escape cultural bias and methods such as analytic induction, constant comparative method and deviant case analysis. Such ideas have led Kalekin-Fishman to say the following (Kalekin-Fishman, 2001, p. 3):

By insisting that these methods as-

> sure "validity", Silverman is defending the positivistic postulate that every type of social science has to be committed to discovering *the* truth, even though throughout the book he defends a constructionist perspective. (Emphasis in the original.)

With towering personalities in qualitative inquiry themselves showing evidence of being conflicted over validity, there is perhaps little hope that minions in the field can be crystal clear of what validity is and is not. Certainly the two writings referred to next could not be of much help either?

First, from Spencer et al. (2003) who conduct a study on quality assessment of qualitative research for the British government, their literature search has led to more than ten kinds of validity – and many of them are interchangeable with other terms. For example, for descriptive validity (Maxwell, 1992; Miller and Fredericks, 1995), the other terms providing the same meaning are descriptive adequacy (Hammersley, 1991) and validity at the individual level (Sykes, 1990). The validity which these terms refer to is concerned with researchers actually capturing what they intended to study and accurately reporting what they have seen or heard. Besides descriptive validity, others pointed out and given the definitions by Spencer et al. (2003, p. 61) include:

- validity of data generation (Mason, 2002) or procedural trustworthiness (Stiles, 1993)
- validity of interpretation (Mason, 2002) or theoretical validity (Maxwell, 1992; Miller and Fredericks, 1995) or explanatory adequacy (Hammersley, 1991)
- interpretive validity (Maxwell, 1992; Miller and Fredericks, 1995)
- validity as incitement to discourse (Lather, 1995)
- reflexive validity (Stiles, 1993) or substantive validation (Angen, 2000)
- dialectic validity (Waterman, 1998)
- critical validity (Waterman, 1998)
- pragmatic validity (Kvale, 1996)
- catalytic or emancipatory validity (Stiles, 1993)

Spencer et al. (2003, pp. 61-62) have also concluded that the different notions of validity in the literature may be categorised into the following areas: method or research process, status of the findings, quality of relations with participants and impact of contribution of the inquiry. This literature finding of theirs is in contrast to their findings from the 29 in-depth interviews with government-based commissioners and managers of research and policy makers, other funders of evaluation research, academics and practitioners involved in conducting qualitative research and writing about quality. Most of the discussion on validity that these interviewees were concerned with is limited to the validity of interpretations and conclusions (Spencer et al., 2003, p. 64). Spencer et al. (2003, pp. 63-64) have also found the existence of different notions of validity together with the different usage of validity terms among the research participants. It seems some research participants can be grouped as the positivists while others, post-positivists.

Besides Spencer et al. (2003), another recent effort which concerns validity criterion is by Onwuegbuzie (2002). To

develop the so-called Qualitative Legitimation Model which attempts to integrate many of the types of validity identified by qualitative researchers, he provides an interesting discussion on validity. He mentions that validity in qualitative research has been operationalized in a myriad of ways with to date no one definition of validity represents a hegemony in qualitative research. The followings are the definitions of validity:

- consensual validity (Eisner, 1992)
- catalytic validity (Lather, 1986)
- ironic legitimation (Lather, 1993)
- paralogical legitimation (Lather, 1993)
- rihizomatic legitimation (Lather, 1993)
- voluptous legitimation (Lather, 1993)
- communicative validity (Kvale, 1995)
- action validity (Kvale, 1995)
- investigation validity (Kvale, 1995)
- descriptive validity (Maxwell, 1992)
- theoretical validity (Maxwell, 1992)
- interpretive validity (Maxwell, 1992)
- evaluative validity (Maxwell, 1992)

Finally, as if the situation over validity is not troublesome enough for the fact that the experts themselves have failed to be consistent and that there appears to be too many facets of validity in qualitative research, another dimension of validity confusion has arisen over the years with many researchers generating or adopting what they consider to be more appropriate terms to describe the qualifying check or measure for their research. So, instead of the term validity, as noted Winter (2000, p. 6), these researchers refer to 'trustworthiness', 'worth', 'relevant', 'plausible', 'confirmable', credible or representative'.

Reliability

Participants in the research conducted by Spencer et al. (2003, p. 65) mentioned earlier view reliability in qualitative research in the following manners: the reassurance that another researcher investigating the same issue or working with the same data set would derive the same findings. But there are researchers who have different views. Stenbacka (2001) argues that since reliability issue relates to measurements it has thus no relevance in qualitative research. Similarly, Golafshani (2003, p. 601) writes that the concept of reliability is irrelevant in qualitative research. He nonetheless next mentions the following (p. 601): "To ensure reliability in qualitative research, examination of trustworthiness is crucial."

The fact that there exist conflicting views over the quality criterion of reliability in qualitative inquiry have however failed to stop Morgan and Drury (2003, p. 6) in detailing out in one long paragraph how qualitative research can attain an appropriate level of research reliability. They write:

> This can be achieved by explaining the methodological framework and the range of strategies that have been used within the study. The rationale for the way in which participants were selected to take part should also be described, as should the researcher's role and their perceived relationship to those participants. It will be necessary to document analytic constructs and meanings, which derive from data, alongside the methodological approach and procedures that were used for producing data. This would include providing de-

> scriptions of phenomena with appropriate narrative of the social context in which they occurred, particularly in terms of persons, places and events. Theoretical propositions also need to be fully explained in terms of how constructs have been formed through detailed procedures.

In short, what they are saying is this: to attain reliability in research, there is a need for the qualitative researcher to document the succession of moves through the stages of data production, analysis and interpretation. It appears however that this explanation by Morgan and Drury (2003) is concerned with the type called external reliability as opposed to internal reliability (Spencer et al. 2003, pp. 64-65).

Indeed, as far as Spencer et al. (2003) are concerned, there are more than one or two notions of reliability. So, besides internal and external reliability which are concerned with the notion of consistency, there are: reliability as replication, inter-coder reliability and auditability, dependability or reflexivity. To achieve the different notions of reliability, there are various steps which a researcher needs to undertake. For internal reliability, for example, the researcher may undertake five different steps including using a team of researchers and peer examination. This is as opposed to the external reliability where the researcher needs to clearly document in the writing another five different matters. These matters include those quoted earlier coming from Morgan and Drury (2003, p. 6).

Generalizability

Typically, the word 'generalizability' is defined as the degree to which the findings can be generalized from the study sample to the entire population (Polit and Hungler, 1991, p. 645 as found in Myers, 2000, p. 2). Since qualitative studies have found it difficult to achieve this, these studies have continued to be criticized for their lack of generalizability. This is despite the many positive aspects of qualitative research and that there are other types of generalizability which qualitative research may still satisfy.

In regard to the latter, Spencer et al. (2003, pp. 67-69) have listed them out as follows:

- representational generalization or generalization within a case (Lewis and Ritchie, 2003);
- analytical or theoretical generalization (Glaser and Strauss, 1967; Strauss and Corbin, 1998);
- empirical or inferential generalization (Stake, 1978; Lincoln and Guba, 1985; Hammersley, 1992).

All in all, just like the criteria of validity and reliability, there is more than one type of qualitative generalization. But it seems to some parties these other types do not exist or are not quite so significance as their favourite one. For them, there is only one kind of generalizability, and it is this very type which qualitative studies would invariably fail to satisfy and which relegates these studies to be among those which they consider to be lacking in rigour.

3. Validity, Reliability and Generalizability in Accounting Research

In order to see the relevance of the quality criteria of validity, reliability and generalizability in accounting research, a total of fifteen papers published in recent years in the top ranked accounting journals is gathered and analysed. See Appendix A. From the outset, it can be easily seen that nearly half or seven of them does not even bother to make any specific reference to these quality criteria. As for the rest, most actually have also failed to say much. The exceptional one is Sweeney and Pierce (2004). But even this one has failed to refer to the quality criterion of generalizability.

Nonetheless, out of the fifteen, aside from Sweeney and Pierce (2004), four – Beattie, Fearnley and Brandt (2004); Gendron, (2002); Gendron and Bedard, (2006); Kwok and Sharp (2005) - manage to provide much details on their section of the research method. This is in contrast to two - Herbach (2005), Rittenberg and Covaleski (2001) - which provide quite minimal discussion on research method. All in all, it is just one paper - Sweeney and Pierce (2004) – which may be said to have attained the level of excellence regarding the quality criteria. Another four papers – Beattie et al. (2004); Gendron (2002); Gendron and Bedard (2006); Kwok and Sharp (2005) – may or may not be at the same level.

In other to find out the truth for these four papers, there is a need for a more detailed analysis over research strategies or practices which a qualitative researcher would normally undertake in order to attain rigour or quality in his or her work and which he or she may disclose in the writing with or without specific reference made to the quality criteria of interest. Such analysis needs first of all the preparation of a table of research strategies and the quality criteria which they satisfy. Next, there is a need for a production of a checklist of appropriate research strategies. Finally, the checklist is used as a basis for comparison with details of research strategies disclosed in those four papers particularly in their research method section. Table 2 provides an example of a set of research strategies which need to be implemented for establishing rigour in qualitative research. This list is prepared based upon discussion in Baxter and Eyles (1997, pp. 506-510).

Once the comparison is made and the result is known, a conclusion may thus be made in regard to the application of any quality criteria. That said, the lack of disclosure of the research strategies for rigour or quality in a research paper or report may not necessarily mean that the study has failed to implement them during the research process. Or, there is the possibility that the researchers and journal editors are those who uphold the quality judgement position that there is *no way* to judge the quality of qualitative studies!

It is just perhaps due to space limitation that leads to their failure in accounting for these research strategies in the research writing. Nonetheless, for the good of everyone involves in qualitative research from authors to editors to readers, it is perhaps worth considering the following coming from Welsh (2002, p. 3):

> Debate on the usefulness of the concepts of validity and reliability

Table 2
Strategies for Attaining Qualitative Rigour

Strategies	Meaning
Rationale for methodology	Qualitative methods are argued to be the most (or only) appropriate way to address the research question
Multiple methods	More than one method used for studying the problem (e.g. in-depth interviews plus participant observation plus textual analysis)
Respondent	A description of the group(s) of respondents (e.g. number and gender ratio is given)
Interview quotations	The words of the respondent may be read or the presentation of verbatim quotations
Interview practices	Details of how interviews were conducted (e.g. use of interview schedules are provided)
Procedures for analysis	A description of how data were converted or condensed into theoretical constructs is given
Immersion of lengthy fieldwork	It is argued that long fieldwork develop rapport with respondents and / or enable deep understandings of the research situation
Revisits	Revisits to respondents are made usually to clarify meanings and build rapport
Verification by respondents	Respondents were contacted to verify interpretations or meanings
Appeals to interpretive community	An existing theory is supported or refuted by the findings, i.e., there is more than reference to the literature
Rationale for verification	Rationale for showing that there is agreement between constructs or interpretations and the meanings held by respondents is provided

> in qualitative research has been undertaken for many years …. Some researchers suggest that whilst these terms are inappropriate in qualitative research, preferring to use terms such as "trustworthiness", "rigorousness", or "quality" of the data, it is nevertheless important that qualitative research and data analysis are carried out in a thorough and transparent manner …. However, in most published research it is unusual to find accounts of exactly how researchers analysed their data and it is partly because of this missing information that this research tradition has been open to allegations of "unthorough" research practices.

That Welsh (2002) is not alone in having such view should not perhaps be surprising. A more recent writing by Sinkovics, Penz and Ghauri (2005, p. 32) seems to point to the same direction:

> Qualitative research methodology … is often criticised for high levels of subjectivity and low reliability and validity. On a substantive level this criticism is unfair because qualitative research offers holistic perspectives on phenomena which cannot be achieved otherwise. However, criticism is often due to a low quality of documentation and reporting of the findings cannot be ignored. While quantitative studies follow a rigorous organisation and presentation in how results are presented, qualitative studies are often reported in a descriptive and narrative way.

4. Discussion and Conclusions

With or without specific references made to quality criteria such as validity, reliability and generalizability in the reporting or publication of a piece of qualitative research, a conscientious researcher who are looking for rigour or quality in their work would be in no doubt see the need for the implementation of the necessary research strategies or practices. Lincoln and Guba (1985), Baxter and Eyles (1997), Morse et al. (2002), Horsburgh (2003) and Shenton (2004) are some of the writings which provide extensive discussion of these research strategies and the quality criteria which they satisfy. In particular, for Baxter and Eyles (1997), their Table II (p. 512) is quite revealing. The same may be said for the chart (p. 73) in Shenton (2004). See Appendix B for a section of this Table II and the chart.

It is also notable that Baxter and Eyles (1997) have provided a list of eight questions which they referred to in the latter half discussion of their paper. It seems an alternative move available in case a list shown in Table 2 earlier appear wanting to some whose qualitative studies use interviews as the main method of data collection. Specifically, it involves answering these eight questions ranging from what was the natural history of the research to what results are presented to how transferable are the findings (Baxter and Eyles, 1997, pp. 511-520). See Appendix C for the full listing of the questions. The following is what Baxter and Eyles (1997, p. 520) say regarding these questions: “It serves not only as a guide for what to look for but reveals where gaps exist in reporting information necessary for ascertaining rigour.” It is also interesting to note that Baxter and Eyles (1997, p. 520) have further proposed that the four trustworthy criteria and the corresponding strategies developed by Lincoln and Guba (1985) be incorporated into the research process as a basis for answering these questions.

Other than Baxter and Eyles (1997), Greenhalgh and Taylor (1997) and Mays and Pope (2000) are other examples of those who produce lists of questions which may be asked over a piece of qualitative research. Greenhalgh and Taylor’s (1997) list of nine questions is however more appropriate for those in medical field. As for Mays and Pope’s (2000) list, check out what they say (p. 52): “We list some questions to ask for any piece of qualitative research (box); the questions emphasise criteria of rele-

vance and validity. They could also be used by researchers at different times during the life of a particular research project to improve its quality." Mays and Pope's list may be found in Appendix C.

Though the presence of checklists should make it easy for researchers to identify which steps to be taken prior to the undertaking of a research or which steps are yet to be implemented while conducting the research, several writers have pointed out the problems that they may create for the same researchers. See Barbour (2001) and Chapple and Rogers (1998) on what these problems are. Notwithstanding these problems, it seems their presence cannot be worse than that of different sets of quality criteria for different types of qualitative studies which some parties have suggested to exist. See Chapple and Rogers (1998), Klein and Myers (1999), Healy and Perry (2000) and Lilford, Edward, Braunholtz, Jackson, Thornton and Hewison (2001). This myriad of quality criteria could very well be one of the reasons leading to researchers failing to make specific references to validity, reliability, generalizability or any other criteria in their writings. And these researchers may include Beattie et al. (2004), Gendron (2002), Gendron and Bedard (2006) and Kwok and Sharp (2005) mentioned earlier. It is just perhaps too troublesome for them to refer to these criteria in their papers when what is important is that they conduct their studies as expected for rigour and that the research strategies performed are reported to the extent possible.

Finally, despite the various positions on quality judgement and that different sets of quality criteria appear to exist for different kinds of qualitative studies, there are around more than a few qualitative studies which can be considered quite enlightening for the fact that they are able to balance the reporting of the story with that of the research process undertaken. In reporting the research process, these papers have also made specific references to quality criteria such as validity, relevance and generalizability to varying degrees. Table 3 provides a list of some of these works and their relevant details.

At least two other types of writing are also available which make specific references to various quality criterion including validity, reliability and generalizability in their discussion. The first is concerned the conduct of qualitative research by students doing PhD. The second concerns the application of specific types of research method for data collection and analysis in qualitative research. Table 4 and Table 5 provide details of the specific examples of the respective types of writings.

All in all, it may safely be said that the qualitative writings listed in Table 3, Table 4 and Table 5 are some of the best examples as far as the incorporation of specific references on quality criteria in the reporting of qualitative research is concerned.

All is apparently not lost!

Table 3
Enlightened Qualitative Studies

Study / Scope	Field	Type of Study	Primary Methods for Data Collection / Data Analysis	Quality Criteria Referred To (number of para.)
Cregan (2005) / Australia	Labour	Induction	Postal Survey / Content Analysis, Hierarchical Cluster Analysis, Textual Analysis	Validity Reliability (several pages!)
Efinger, Maldonado, McArdle (2004) / United States	PhD courses	Phenomenology - Constructivist	Structured, Open-Ended Questionnaire / Content Analysis	3 Trustworthiness* 5 Authenticity** (5)
Waldman, Lituchy, Gopalakrishnan, Laframboise, Galperin and Kaltsounakis (1998) / United States and Canada	Quality	Multiple Case-Study	Open-Ended Interviews/ Pattern-Matching	Accuracy Objectivity Reliability Validity (3)
Riley (1995) / United States	Tourism	Situationalism	Long Interviews / Grounded Theory	Credibility, Dependability, Confirmability (3)

Note: * Dependability, Tranferability and Credibility
** Fairness, Ontological Authenticity, Educative Authencity, Catalytic Authencity and Technical Authencity

Table 4
The Conduct of Qualitative Research for a PhD

Study / Scope	Research Strategy	Primary Methods for Data Collection / Analysis	Quality Criteria Referred To (number of para.)
Bowen (2005) / United States	Exploratory study	In-Depth, Open-Ended Interviews / Grounded Theory	4 Trustworthiness (7)
Peterson and Higgs (2005) / Transnational	Hermeneutics	Conversational Interviews / Hermeneutics	Credibility* Rigor Ethical (10)
de Weerd-Nederhot (2001) / The Netherlands	Case Study	Interviews/Within – And Cross – Case Analysis *ala* Miles and Huberman (1994)	Variety! (several pages!)
McCotter (2001) / United States	Deconstruction	Variety	Validity (4)
Carter (1999) / Canada	Multiple case study	Semi-Structured Interviews / Grounded Theory	4 Trustworthiness (5)

Note: *Authenticity, Plausibility, Trustworthiness

Table 5
The Use of Specific Research Methods for Data Collection / Analysis in Qualitative Studies

Study / Scope	Field of Study	Primary Methods for Data Collection / Analysis	Quality Criteria Referred To (number of para.)
Rao and Perry (2003) / Australia	Marketing/ Internet	Convergent Interviews	Construct Validity Internal Validity External Validity Reliability (7)
Callahan and Elliot (1996) / United States	Behavioural Economics	Free Narrative	Variety! (several pages!)
Yeung (1995) / Hong Kong and ASEAN	International Business	Qualitative Personal Interviews	Validity Reliability (several pages!)

References

Adams, C.A., & Harte, G. (1998). The changing portrayal of the employment of women in British banks' and retail companies' corporate annual reports. *Accounting Organizations and Society, 23*(8), 781-812

Ahrens, T., & Chapman, C.S. (2006). Doing qualitative field research in management accounting: Positioning data to contribute to theory. *Accounting, Organizations and Society, 31*, 819-841.

Anderson-Gough, F., Grey, C., & Robson, K. (2001). Tests of time: organizational time-reckoning and the making of accountants in two multi-national accounting firms. *Accounting, Organizations and Society*, *26*(2), 99-122.

_____ (2005). "Helping them to forget..": The organizational embedding of gender relations in public audit firms. *Accounting, Organizations and Society, 30*(5), 469-490.

Angen, M.J. (2000). Evaluating interpretive inquiry: Reviewing the validity debate and opening the dialogue. *Qualitative Health Research, 10*(3), 378-395.

Arena, M., Arnaboldi, M., & Azzone, G. (2006). Internal audit in Italian organizations: A multiple case study. *Managerial Auditing Journal*, *21*(3), 275-292.

Atkinson, A.A., & *Shaffir, W.* (1998). Standards for field research in management accounting. *Journal of Management Accounting Research*, *10*, 41-68.

Barbour, R.S. (2001). Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *British Medical Journal, 322*, 1115-1117.

Barusch, A., Gringeri, C., & George, M. *(2011). Rigor in qualitative social work research: A review of strategies used in published articles. Retrieved June 02, 2011 from http://www.faqs.org/ periodicals/201103/2302649611.html#ixzz1Nl0kfxpM*

Baxter. J., & Eyles, J. (1997). Evaluating qualitative research in social geography: Establishing 'rigour' in interview analysis. *Transactions of the Institute of British Geographers, 22*(4)*,* 505-525.

Beattie, V., Fearnley, S., & Brandt, R. (2004). A grounded theory model of auditor-client negotiations. *International Journal of Auditing*, *8* (1), 1-19.

_____ (2001). *Behind closed doors: What company audit is really about.* Basingstoke, Hampshire: Palgrave.

_____ (2000). "Behind the audit report: a descriptive study of discussion and negotiations between auditors and directors", *International Journal of Auditing*, *4*(2), 177-202.

Bergman, M.M., & Coxon, A.P.M. *(2005*). The quality in qualitative methods. *Forum Qualitative Social Research [On-line Journal], 6*(2). Retrieved September 03, 2006, from http://www.qualitative-research.net/fqs-texte/2-05/05-2-34-e.htm

Bowen, G.A. (2005). Preparing a qualitative research-based dissertation: Lessons learned. *The Qualitative Report, 10*(2), 208-222. Retrieved September 03, 2006, from http://www.nova.edu/ssss/QR/QR10-2/bowen.pdf

Brown, S.M. (1992). Cognitive mapping and repertory grids for qualitative survey research: Some comparative observations. *Journal of Management Studies*, *29*(3), 287-307.

Callahan, C., & Elliot, C.S. (1996). Listening: A narrative approach to everyday understandings and behavior. *Journal of Economic Psychology, 17*(1), 79-114.

Carter, M. (1999). *A profile of service-learning programs in South Carolina and their responsiveness to the national priorities*. Columbus, Ohio: Bell & Howell Company.

Chapple, A., & Rogers, A. (1998). Explicit guidelines for qualitative research: A step in the right direction, a defence of the 'soft option' or a form of sociological imperialism? *Family Practice*, *15*, 556-561.

Cregan, C. (2005). Can organizing work? An inductive analysis of individual attitudes toward union membership. *Industrial and Labour Relations Review 58*(2), 282-304.

Creswell, J.W. (2007). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publications.

Creswell, J.W., & Miller, D.L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*(3), 124-131.

Davies, D., & Dodd, J. (2002). Qualitative research and the question of rigor. *Qualitative Health Research, 12*, 279-289.

de Weerd-Nederhot, P.C. (2001). Qualitative case study research: The case of a PhD research project on organising and managing new product development systems. *Management Decision, 39(*7), 513-538.

Denzin, N.K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York: McGraw-Hill.

_____ (1995). The experiential text and the limits of visual understanding. Educational Theory, 45 (1), 7-18.

_____ & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research.* Thousand Oaks, CA: Sage Publications.

Eberle, T.S. (2005). Promoting qualitative research in Switzerland. *Forum Qualitative Social Research [On-line Journal*], *6*(2). Retrieved September 03, 2006, from http://www.qualitative-research.net/fqs-texte/2-05/05-2-31-e.htm

Efinger, J., Maldonado, J., & McArdle, G. (2004). PhD students' perceptions of the relationship between philosophy and research: A qualitative investigation. *The Qualitative Report, 9*(4), 732-759. Retrieved September 03, 2006, from http://www.nova.edu/ssss/QR/QR9-4/efinger.pdf

Eisenhardt, K.M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532-550.

Eisner, E. (1992). Objectivity in educational research. *Curriculum Inquiry*, *22(*1), 9-15.

Gendron, Y. (2002). On the role of the organization in auditors' client-acceptance decisions. *Accounting, Organizations and Society, 27*(7), 659-684.

_____ , & Bedard, J. (2006). On the constitution of audit committee effectiveness. *Accounting, Or-*

ganizations and Society, 31(3), 211-239.

_____ ., Cooper, D.J., & Townley, B. (2007). The construction of auditing expertise in measuring government performance. *Accounting, Organizations and Society, 32*(1/2), 101-129.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies of qualitative research*. London: Wiedenfeld and Nicholson.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597-607.

Greenhalgh, T. & Taylor, R. *(*1997). *How to read a paper: Papers that go beyond numbers (qualitative research). British Medical Journal,* 315(7110), 740.-743.

Grey, C., Robson, K., & Anderson, F. (1997). *Professional socialization in chartered accountancy practices: Making up accountants*. London: ICAEW.

Guba, E.G., & Lincoln, Y.S. (1981). *Effective evaluation*. San Francisco: Jossey-Bass Publishers.

Guba, E.G., & Lincoln, Y.S. (1994). *Handbook of qualitative research.* Thousand Oaks, CA: Sage Publications.

Hammersley, M. (1991). *Reading ethnographic research*. London: Longman.

_____ (1992). *What's wrong with ethnography?* London: Routledge.

Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research*, *3*(3), 118-126.

Herrbach, O. (2001). Audit quality, auditor behaviour and the psychological contract. *European Accounting Review*, *10*(4), 787–802.

Horsburgh, D. (2003). *Evaluation of qualitative research. Journal of Clinical Nursing, 12(2), 307-312.*

Huberman, M., & Miles, M.B. (1991). *Analyse des données qualitatives: Recueil de nouvelles Methods*. Brussels: De Boeck-Wesmael.

Irvine, H.J., & Gaffikin, M. (2006). Getting in, getting on and getting out: Reflections on a qualitative research project. *University of Wollongong Research Online.* Retrieved June 02, 2011 http:// ro.uow.edu.au/cgi/ viewcontent.cgi? article=1127&context= commpapers&sei-redir=1#search=% 22Irvine+H,+Gaffikin+M,+% 282006% 29+Getting+in,+getting+on+and+ getting+ out:%22

Janesick, V.J. (1994). The dance of qualitative research design: Metaphor, methodolatry, and meaning. In Denzin, N.K. & Lincoln, Y.S. (Eds.), *Handbook of qualitative research* (pp. 209-235). Thousand Oaks, CA: Sage Publications.

Kalekin-Fishman, D. (2001). Review: David Silverman (2001). Interpreting qualitative data: Methods for analysing talk, text and interaction. *Forum Qualitative Social Research, 2*(3). Retrieved September 03, 2006, from http://nbn-resolving.de/urn:nbn:de:0114-fqs010363

King, N. (1994). The qualitative research interview. In Cassell, C., & Symon, G. (Eds.), *Qualitative methods in organizational research: A practical guide* (pp. 14-

36). London: Sage Publications.

Klein, H.K., & Myers, M.D. (1999). Investigating information systems with ethnographic *research. Communications of the Association for Information System*, *2*(23), 1-20.

Kuzel, A. & Engel, J. (2001). Some pragmatic thought on evaluating qualitative health research. In Morse, J., Swanson, J., & Kuzel, A. (Eds.). *The nature of qualitative evidence* (pp. 114-138). Thousand Oaks, CA: Sage Publications.

Kvale, S. (1989). *Issues of validity in qualitative research.* Lund, Sweden: Chartwell Bratt.

_____ (1995). The social construction of validity. *Qualitative Inquiry*, *1*, 19-40.

_____ (1996). *Interviews: An introduction to qualitative research.* Thousands Oaks, London, New Delhi: Sage Publications.

Kwok, Chee Chiu, & Sharp, D. (2005). Power and international accounting standard setting: Evidence from segment reporting and intangible assets projects. *Accounting, Auditing and Accountability Journal*, *18*(1), 74-99.

Lather, P. (1986). Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange*, *17*(4), 63-84.

_____ (1993). Fertile obsession: Validity after poststructuralism. *The Sociological Quarterly*, *34*(4), 673 -93.

_____ (1995). Post-critical pedagogies: A feminist reading. In McLaren, P. (Ed.) *Postmodernism, postcolonialism and pedagogy* (pp. 167-186). Albert Park: James Nicholas Publishers.

Lazaraton, A. (2003). Evaluative criteria for qualitative research in applied linguistics: Whose criteria and whose research? *Modern Language Journal, 87*, 1-12.

Lewis, J., & Ritchie, J. (Eds.). (2003). *Qualitative research practice - A guide for social science students and researchers*. London: Sage Publications.

Lilford, R.J., Edwards, S., Braunholtz, D.A., Jackson, J., Thornton, J., & Hewison, J. (2001). Ethical issues in the design and conduct of randomised controlled trials. In Stevens, A.J.H., Abrams, K., Brazier, J., Fitzpatrick, R., Lilford, R.J. (Eds.). *Advanced Handbook of Methods in Evidenced Based Healthcare* (pp. 11-24). Thousand Oaks, CA: Sage Publications.

Lillis, A.M (2006). Reliability and validity in field study research. In Hoque, Z. (Ed.) *Methodological issues in accounting research: Theories and methods* (pp. 461-475). London: Spiramus.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage Publications.

Mason, O.J. (2002). Teaching qualitative research methods: Some innovations and reflections on practice. *Psychology Teaching Review, 10* (1), 68-75.

Maxwell*, J.A.* (1992). Understanding and validity in qualitative research. *Harvard Educational Review, 62*(3), 279-300.

_____ (2004). *Qualitative research design: An interactive approach (*2nd ed.). Beverly Hills, CA: Sage Publications.

Mays, N. & Pope, C. (2000). Qualitative research in health care: Assessing quality in qualitative research. *British Medical Jour-*

nal. *320*(7226), 50-52.

McCotter, S.S. (2001). The journey of a beginning researcher. The *Qualitative Report, 6*(2). Retrieved September 03, 2006, from http://www.nova.edu/ssss/QR/QR6-2/*mccotter*.html

McKinnon, J. (1988). Reliability and validity in field research: some strategies and tactics. *Accounting, Auditing and Accountability*, *1*, 34-54.

Miles, M.B. & Huberman, A.M. (1994), *Qualitative data analysis: An expanded sourcebook* (2nd ed.). London: Sage Publications.

Miller, S., & *Fredericks, M*. (1995). The nature of 'evidence' in qualitative research methods. *Qualitative Health Research*, *5,* 388-397.

Modell, S. (2005). Triangulation between case study and survey methods in management accounting research: An assessment of validity implications. *Management Accounting Research*, *16*, 231-254.

Morgan, *A.K., & Drury, V.B. (2003*). Legitimising the subjectivity of human reality through qualitative research method. *The Qualitative Report*, *8*(1). Retrieved September 03, 2006, from http://www.nova.edu/ssss/QR/QR8-1/morgan.html

Morse J.M., Barrett M., Mayan M., Olson K. & Spiers J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 1*(2), 1-19.

Myers, M. (2000). *Qualitative research and the generalizability question: Standing firm with proteus. The Qualitative Report,* 4(3/4). Retrieved September 03, 2006, from http://www.nova.edu/ssss/QR/QR4-3/myers.html

Neimark, *M.D.* (1992). *The hidden dimensions of annual reports*. London: Paul Chapman.

Onwuegbuzie, A.J. (2002). Positivists, post-positivists, post-structuralists, and post-modernists: Why can't we all get along? Towards a framework for unifying research paradigms. *Education, 122*(3), 518–530.

Patton, M.Q. (1990), *Qualitative evaluation and research methods* (2nd ed.). Newbury Park: Sage Publications.

Peterson, M., & Higgs, J. (2005). Using hermeneutics as a qualitative research approach in professional practice. *The Qualitative Report*, *10*(2), 339-357.

Polit, D., & Hungler, B. (1991). *Nursing research: Principles and methods*. New York: JB Lippincott.

Rao, S., & Perry, C. (2003). Convergent interviewing to build a theory in under-researched areas: Principles and an example investigation of internet usage in inter-firm relationships. *Qualitative Market Research: An International Journal*, *6(*4), 236-247.

Riley, M. (1995). Tourism development under close control: The case of the Falkland Islands. *Tourism Management*, *16*(6), 471-474.

Rittenberg, L.E., & Covaleski, M.A. (2001). Internalization versus externalization of the internal audit function: An examination of professional and organizational imperatives. *Accounting, Organizations and Society 26*, 617–641.

Rubin, H., & Rubin, I. (1995). *Qualita-*

tive interviewing: The art of hearing data. London: Sage Publications.

Sarens, G., & De Beelde, I. (2006), Building a research model for internal auditing: Insights from literature and theory specification cases. *International Journal of Accounting, Auditing and Performance Evaluation*, *3*(4), 452-470.

Seale, C. (1999). Quality in qualitative research. *Qualitative Inquiry, 5* (4), 465-478.

Shenton, A.K. (2004). The analysis of qualitative data in LIS research projects: A possible approach. *Education for Information, 22*, 143-162.

Silverman, D. (2001). *Interpreting qualitative data: Methods for analysing talk, text and interaction* (2nd ed.). London: Sage Publications.

Sinkovics, R.R., Penz, E., & Ghauri, P.N., (2005). Analysing textual data in international marketing research. *Qualitative Marketing Research: An International Journal*, *8*(1), 9–38.

Siti-Nabiha, A.K., & Scapens, R.W. (2005). Stability and change: An institutionalist study of management accounting change. *Accounting, Auditing and Accountability Journal*, *18*(1), 44-73.

Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). *Quality in qualitative evaluation: A framework for assessing research evidence*. London: National Centre for Social Research.

Stake, R. (1978). The case study method in social inquiry. *Educational Researcher, 7*(2), 5-8.

Stenbacka, C. (2001). Qualitative research requires quality concepts of its own. *Management Decision, 39*(7), 551-555.

Stiles, W.B. (1993). Quality control in qualitative research. *Clinical Psychology Review, 13*, 593–618.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage Publications.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). London: Sage Publications.

Suter, E.A. (2010). Validity in qualitative research on personal relationships. *Kentucky Journal of Communication* special issue on the future of relational communication. Retrieved June 02, 2011 from https://portfolio.du.edu/pc/port?portfolio=esuter

Sweeney, B., & Pierce, B. (2004). Management control in audit firms: A qualitative examination. *Accounting, Auditing and Accountability Journal*, *17(*5), 779-812.

Sykes, W. (1990).Validity and reliability in qualitative market research: A review of the literature. *Journal of the Market Research Society*, *32* (3), 289-328.

Thompson, *J.B.* (1988). *Critical Hermeneutics* (4th ed.). Cambridge: Cambridge University Press.

Unerman, J. (2000). Reflections on quantification in corporate social reporting content analysis. *Accounting, Auditing and Accountability Journal, 13*(5), 667-680.

Unerman, J. (2003). Enhancing organizational global hegemony with narrative accounting disclosures:

An early example. *Accounting Forum, 27*(4), 425-448.

Van Peursem, K.A. (2004). Internal auditors' role and authority: New Zealand evidence. *Managerial Auditing Journal*, *19*(3), 378-393.

_____ (2005). Conversations with internal auditors: The power of ambiguity. *Managerial Auditing Journal*, *20* (5), 489-512.

Waldman, D.A., Lituchy, T., Gopalakrishnan, M., Laframboise, K., Galperin, B., and Kaltsounakis, Z. (1998). A qualitative analysis of leadership and quality improvement. *Leadership Quarterly*, *9*(2), 177-201.

Waterman, H. (1998). Embracing ambiguities and valuing ourselves: Issues of validity in action research. *Journal of Advanced Nursing, 28* (1), 101-105.

Welsh, E. (2002). Dealing with data: Using NVivo in the qualitative data analysis process. *FQS, 3*(2). Retrieved September 03, 2006, from http://www.qualitative-research.net/fqs-texte/2-02/2-02 welsh-e.htm

Werner, O., & Schoepfle, G. (1987). *Systematic Fieldwork: Vol.1, Foundations of Ethnography and Interviewing*. Newbury Park, CA: Sage.

Whittemore, R., Chase, S.K., & Mandle, C.L. (2001). Validity in qualitative research. *Qualitative Health Research*, *11*, 522-537.

Winter, G. (2000). A comparative discussion of the notion of 'validity' in qualitative and quantitative research. *Qualitative Report*, *4*(3/4). Retrieved September 03, 2006, from http://www.nova. edu/ssss/ QR/QR4-3/winter.html

Wolcott, H. (1994). *Transforming qualitative data: Description analysis and interpretation.* Thousand Oaks, CA: Sage Publications.

Yeung, H. Wai-chung. (1995). Qualitative personal interviews in international business research: Some lessons from a study of Hong Kong Transnational Corporations. *International Business Review, 4* (3), 313-339.

Yin, R. (1994). *Case study research: Design and method* (2nd ed.). Beverly Hills, CA: Sage Publications.

APPENDIX A
Validity, Reliability and Generalizability in Accounting Research

APPENDIX A

Research	Journal	Research Method	Validity, Reliability, Generalizability, Etc.
Anderson-Gough, Grey and Robson (2001)	AOS	Qualitative methodology based principally upon a programme of face-to-face semi-structured interviews with audit trainees employed by accountancy practices. Interviews took place in a private room on the premises of the employing firm. Interviewees were all undertaking their professional training contract (examinations) with the Institute of Chartered Accountants in England and Wales (ICAEW). • Although the interview programme was the principal research instrument employed, there are three other sources of material. • The interviews fall in two parts. First, 77 interviews were conducted with ICAEW trainees in two offices of two Big Six firms located in cities in the North of England for a project on professional socialisation (Grey, Robson and Anderson, 1997). Second, around 30 interviews were conducted with newly qualified seniors and managers of the same two firms as part of a project concerned with socialisation and career progression in audit firms. Some of the interviewees had been interviewed a few years earlier while they were still trainees. Also, some exit interviews were conducted on staff leaving or who had left the two firms. The first series of interviews took place between January 1996 and March 1997. The second during 1999. • Where permission was given, interviews were tape-recorded. • Recordings of interviews were transcribed and coded for *Ethnograph*, a qualitative data analysis programme. (There is also an explanation of the coding work in two rather long paragraphs.)	

Sweeney and Pierce (2004)	AAAJ	A qualitative approach using 25 semi-structured interviews (of approximately 50 minutes each in an off-site setting) of auditors in their third and fourth years of employment in four of the (then) Big Five firms in Ireland in 2000. *Note*: Though the following is said on the second page of the paper – "The purpose of this study is to develop a deeper understanding of control system variables previously shown to be related to the incidence of QTB [quality-threatening behaviour] in audit firms …"- in the concluding section, it is mentioned: The study is exploratory in nature …" • The data collection was carried out by the first author and both authors analysed the data. • The number of interviewees was determined during data collection at the point where no new insights were being obtained from interviewees and saturation (Lincoln and Guba, 1985) was reached. (However, elsewhere it is noted that approximately equal numbers of third and fourth year auditors were selected by the researchers at random from staff listings for interview, reflecting a roughly even gender balance and the presence of auditors from each of the main audit specialism.) • Demographic details of interviewees' gender, firm, length of experience in years and audit specialism is made available. • Interview guide approach is used with sections of the guide related to the paper reproduced in appendix. • On interviewing, there is assurance of confidentiality and permission sought from interviewees for recording of the interview.	p. 787: "While recognising that no single research method can be truly objective, it is important to document all the steps taken to increase *objectivity* as far as possible and to be constantly alert for *subjectivity* at data collection and analysis stages (Patton, 1990). Previous field research has been criticised for failing to attend to such research criteria as *validity* and *reliability* (McKinnon, 1988). Several steps were taken to limit *bias* and increase *objectivity* both during the interview and in analysing the interview data." (Emphasis added.) (These steps which they explained right after together with their rationale involve the followings: the use of interview guide; the taking of notes during interviews; the reviewing of transcripts of the first two interviews by a colleague and the use of structured analytical method in analysis of data.) p. 788: "In analysing the data, the researchers read fully through each of the coded transcripts three times … To help ensure a *rigorous*, *complete* and *impartial* analysis of the findings (Lillis, 1999) an interviewee-specific summary of key findings was prepared and this is set out in Table II … As recommended by Miles and Huberman (1994), some quantitative analysis of our findings was carried out using frequency counts to test possible *bias* and to get an indication of the *robustness* of the findings." (Emphasis added.)

		• All interviewees gave permission to tape the interview and no interviewee asked for the tape to be turned off at any stage. All interviewees seemed relax and were willing to answer questions honestly. • Following each interview, transcribing took place in separate Microsoft Word document. Next, a contact summary sheet is prepared consisting of brief answers to each of questions on interview schedule. Finally, the transcript is coded involving the importing of the Word document into NUDIST. All this is done before the next interview so that the researcher may learn from the previous interview.	p. 788: "For presentation of findings, sentences which appeared to represent a particular code/ theme were used to present the 'thick description' (Denzin, 1994, p. 505) in the findings section … Regarding the *quality* of the findings presented, the researchers checked to ensure there were no contradictory statements made by the same interviewee and that the evidence presented to back up each finding appeared to be *reliable*. Following the first draft of the 'thick description' prepared from individual reports of each code, each of the transcripts was fully read again to ensure that each of the quotations selected was considered in context and further revisions were made in the presentation of findings." (Emphasis added.) p. 807: "The preceding discussion needs to be considered in the context of the strengths and limitations of the study. A particular concern was the possible existence of various forms of *bias* inherent in qualitative research and careful attention was paid to the pursuit of *rigorous* and *comprehensive* approach to collection and analysis of data, as reported earlier. The findings can therefore be viewed as having a high degree of *internal validity* (Patton, 1990)." (Emphasis added.)
Herbach (2005)	AAAJ	15 semi-directed interviews where 13 were conducted with volunteered audit seniors who earlier were respondents (from a total of 135) of a questionnaire research on audit quality reduction (AQR) behaviours (Herrbach, 2001). These auditors were working in the French office of four Big Five. Another two interviews took place with two junior audit managers who had participated in the questionnaire's pre-test. • To gain the group of 13 audit seniors as interviewees, a summary of results of the questionnaire study was sent to all 135 respondents and the cover letter proposed that they take part in interviews to gain a deeper understanding of the phenomenon.	p. 392: "… due to the number of interviews (15) and the situated context (French audit seniors and former seniors), its *generalisability* is limited and it should be viewed as an exploratory study whose insights are used to provide recommendations for future research into various aspects of auditor professionalism." (Emphasis added.)

		• The research instrument is placed in the appendix of the paper. • Fifteen interviewees were considered sufficient for this study, because little new or relevant data seemed to emerge from the latter interviews, which tends to imply that the interview results were becoming 'theoretically saturated' (Strauss and Corbin, 1990).	p. 402: "… there are also obvious *generalisation* problems. The interviews were performed with current and former French Big Five auditors and at one point in time, which raises the question of the *validity* of the findings for other ranks in audit firms, other firms than the (then) Big Five, other cultures than the French and other times than the turn of the millennium. The potential impact of the French national context, in particular, has to be kept in mind." (Emphasis added.)
Beattie, Fearnley and Brandt (2004)	IJA	Matched, in-depth interviews conducted with audit partners and finance directors of six major UK listed companies who had recently experienced interactions involving significant accounting issues. A total of eleven interviews recorded and fully transcribed. For the remaining interview, extensive notes were taken and dictated immediately afterwards. *Note*: The paper is a summary of a qualitative case study earlier reported in a book (Beattie, Fearnly and Brandt, 2001). On page 3, it is stated: "This makes the findings accessible through the journal literature." Within the book, there is a detailed coverage of the development of grounded theory. On page 3 too, it is next stated: "This permits the 'huge chasm' that often separates data from conclusions to be bridged." • Identification of six cases came through from Beattie, Fearnly and Brandt (2000) questionnaire study. Six finance directors (FDs) who indicated high levels of negotiation and discussion were asked whether they were willing to be interviewed to enable an exploration of their responses in greater depth. They all agreed. Eisenhardt (1989) recommended four to ten optimum number of cases in this type of analysis. Cases represent a range of companies sizes, industry sectors and audit firms. • Written assurances given that neither the interviewee nor the company would be identified or identifiable in subsequent publication. • No standard interview guidelines used. However, prior to interviews, company's annual reports for the period covered by the questionnaire were studied as familiarisation exercise.	p. 17, footnote 2: "To support the accuracy of the interviewee's statements, reference was made to the company's annual reports where the outcomes of some of the discussions and negotiations were observable, thus providing assurance about the *reliability* of the evidence collected (Yin, 1984, p. 80)." (Emphasis added.) (This footnote is attached to the line on page 6 which says that at the end of each interview, the FD was asked for permission to interview the AEP.)

		• Application of neutral, conversational prompts and a laddering technique during interviews. This means the interviewer keeps asking 'why?', working backwards to antecedent conditions and forwards to anticipated effects (Brown, 1992, p. 293). • Also, during interviews, FDs were asked 'to tell the story' from their perspective of the discussions and negotiations with their auditors referred to in their questionnaire responses. They were also encouraged to raise any other issues they wanted to (Thompson, 1988). • At the end of each interview, the FD was asked for permission to interview the audit engagement partner (AEP) with whom the discussions and negotiations had taken place. All the FDs gave their consent. During interviews, each AEP was asked to 'tell the story' from his perspective about the issues discussed with the clients. He was also encouraged to add any other information he considered relevant. • Analysis of transcribed interviews used grounded theory procedures and techniques (Strauss and Corbin, 1990). There is an extensive explanation of the four stage process involving three types of coding types (open, axial and selective) and their outputs in the form of variety of concepts, categories and relationships.
Anderson-Gough, Grey, Robson (2005)	AOS	Qualitative methodology based principally upon a programme of face-to-face semi-structured interviews with audit trainees employed by accountancy practices, and conducted face-to-face in a private room on the premises of the employing firm. Interviewees were all undertaking their professional training contract (examinations) with the Institute of Chartered Accountants in England and Wales (ICAEW). *Note*: On page 2, it is stated that the paper is based upon two intensive projects of socialisation and professional identity in offices of two Big Five firms in UK, "… studies which we believe to be the largest of their kind in this country." Interviews were mentioned to have been conducted from 1995 to 2000. • Although the interview programme was the principal research instrument employed, there are three other sources of material. • The interviews fall in two parts. First, 77 interviews were conducted with ICAEW trainees in two offices of two Big Six firms located in cities in the North of England for a project on professional socialisation (Grey, Robson and Anderson, 1997). Second, around 30 interviews were conducted with newly qualified seniors and managers of the same two firms as part of a project concerned with socialisation and career progression in audit firms. Some of the interviewees had been interviewed a few years earlier while they were still trainees. Also, some exit interviews were conducted on staff leaving or who had left the two firms. The first series of interviews took place between January 1996 and March 1997. The second during 1998 and 1999.

		• Recordings of interviews were transcribed and coded for *Ethnograph*, a qualitative data analysis programme. (There is also an explanation of the coding work in two rather long paragraphs.)	
Gendron (2002)	AOS	A field study at three Big Six Canadian firms. But the paper reports data collected from two of the firms due to space limitations. That said, data from this particular firm which is like that of one of the two other firms gives support to the study's theoretical adaptation. • Fieldwork and analysis used ideal types. The two ideal types were developed a priori by relying on auditing literature. These ideal types, in turn were used to develop interview instruments, and provided benchmarks to which empirical data were compared. • Sources of information used in each firm to gather data specified in a tabular form. Clearly, data came mainly from semi-structured interviews conducted mainly in offices located in Montreal and Toronto. There were in a total seven interviews conducted in one firm and another six for the other firms. Most interviewees were partners. In each firm, "to mitigate selection-bias risk", the contact partners were asked to provide interviewees that varied in their length of service with the firm. Other information sources were firms' internal documents and descriptive brochures. • Various "tactics to reduce response bias and overcame interviewees' potential reluctance to freely provide information" were incorporated in the interviews. This included not taping of the interviews "… since the presence of a tape-recorder may have induced interviewees not to provide certain information." • Data was analysed by relying on qualitative procedures (Huberman and Miles, 1991; Patton, 1990). Next, it is stated: "Descriptions of decisions and other documents were coded to identify the main theme(s) of each segment of data. Tables and figures were developed to display data in a compressed and ordered form. Conclusions emerged from comparisons of decisions and sites, and were verified by searching for contradictory data." • In one long paragraph, details of the qualitative analytical process were provided beginning with the sentence "Specifically, qualitative procedures began during fieldwork." • After site reports were completed, each was sent to the corresponding contact partner. Each contact partner was given a month to communicate any concerns to the researcher.	p. 680: "… this study is inevitably characterised by limitations. First … Second, the findings are inescapably reliant on the quality and comprehensiveness of the data collected. In this respect, it is worth underlining that data collection was limited due to the firms' participation requirements. The firms did not allow me to interview all decision-makers involved in each client-acceptance situation, nor to examine the corresponding documentation. The descriptions of decisional situations therefore are unlikely to be as comprehensive as they could have been. Nonetheless, several steps were taken to increase the *reliability* of the field study. Complete anonymity was provided to the firm and the interviewees who, furthermore, were not made aware that the professional and commercial logics of action were being investigated." (Emphasis added.) p. 680: "… since data collection was completed in Big Six offices located in the two largest cities of Canada, readers should be careful before *generalizing* the results to other settings. For example, in Big Six offices located in smaller cities, signals from the firm's organizational components may affect auditors' mindset differently …"

Rittenberg and Covaleski (2001)	AOS	Qualitative analysis of archival material (both public and private records as well as business press coverage of the events examined) supplemented with extensive interviews with key parties "to elicit their views and their guidance on accessing relevant material." • Latent or qualitative content analysis (p. 624): "… the researcher serves as a research instrument in interpreting archival material (Van Maanen, 1979, 1988)."	
Sarens and De Beelde (2006)	MAJ	Qualitative and descriptive study involving relevant documents and chief audit executives from a total of ten manufacturing and service companies: four Belgian subsidiaries of US companies and the rest Belgian companies. Each company had at least 1000 employees. • The selection of companies was theoretically driven and not by a concern of representativeness since the aim is never to generalise the findings to other settings. Two specific criteria were applied in making the selection. • QualitaIn-depth interviews took place with the chief audit executive in each of the ten companies. Six preliminary interviews formed a basis to develop a more structured and focused interview guide for the in-depth interviews. (There is a paragraph disclosing the four parts of the interview.) Each interview lasted from 60 to 90 minute, was tape recorded and transcribed immediately after the interview took place. In the endnote no. 2 attached to this very detail, the following is stated: "These steps in the data collection procedure were performed by research assistants. They had the necessary background on the research topic … A closed interview guide was used in order to minimize interviewer bias …" • There is a statement on triangulation on p. 68: "In order to triangulate the interview data, we used archival data, like the internal audit charter, the audit committee charter … obtained from the interviewees."	p. 70: "While interview data may enhance *construct validity* by studying phenomena in their natural context, it is suggested by Lillis (1999) to use a systematic analytical protocol to enhance the *reliability* of our results. More specifically, we referred to the most important steps from the analytical protocol suggested by Miles and Huberman (1994). First, all interview transcripts and archival documents were coded. Next, we structured and summarized the insights of each company in order to get an overview of the most remarkable insights … The 'translation' of these insights into a standardized and comparable matrix was an important tool to facilitate cross company analysis. Finally, we compared the ten companies in order to discover certain patterns and reassure ourselves that conclusions from one company were not idiosyncratic." (Emphasis added.) p. 78: "Given the qualitative nature of this study and the limited number of companies, it is important to take into account that *generalization* of the results to all Belgium and US companies is not possible." (Emphasis added.)

Van Peursem (2005)	MAJ	A qualitative approach - specifically methods used by Eisenhardt (1989) for frame-braking and theory development. • Six cases comprising of organisations which satisfy Eisenhardt's (1989) idea of cases defining particular domain and filling theoretical categories. • Those interviewed coming from the six selected organisations were earlier involved in a survey conducted by the same researcher (Van Peursem, 2004). In the paper's appendix, there is a list detailing the six interviewees' backgrounds and the sites where the interviews had taken place. • Coding procedures drew from Eisenhardt's suggestions. For the within-case analysis, prior to the discussion of the findings, there is a rather lengthy discussion of what actually it entails. • For the purpose of analysis, early on in the paper, the following is stated (p. 491): "To add further rigour to the process, a independent researcher entered the analysis once a tentative structure had been posed, and he examined the original data …"	p. 491: "Each internal auditor was interviewed, and their working environment was observed. Document, where raised in conversation, were examined and discussed. Some confirmatory actions were taken to ensure a clear understanding of their intent, and this included following up on their points during the course of the interviews (see Connell *et al*., 2001, on *internal validity* issues and Eisenhardt, 1989, on ideas of overlap and opportunistic data collection). Once the interviews were transcribed, the raw data was fed back to them for further comment or elaboration." (Emphasis added.)
Arena, Arnaboldi and Azzone (2006)	MAJ	• A multiple case study comprising of six companies. • Data collected through semi-structured interviews with organisation personnel and documentation. It is noted (p. 282): "We tried to enhance the research by using multiple sources of information and verifying this when it was possible with the data that emerged from interviews." • The theoretical framework leads to the findings. As stated out (p. 282): "In accordance with the framework we identified three categories of companies with different features of Internal Audit Departments that we expected to find: companies in which no internal audit structure has been introduced …" • For analysis, it is simply stated that the three categories of companies are analysed on the basis of three aspects (internal audit department characteristics; activities performed by internal auditors …) and that for each company there is the specification of the regulations with which it has to comply.	p. 282: "Despite concerns that case studies of specific situations do not afford results which may be regarded as generally applicable or *generalizable*, the case study method has been selected for this investigation because of its potential to provide a richer, more detailed understanding of internal audit practices. The case study method also offered the possibility of achieving triangulation (Denzin, 1978) by extending beyond the formal interviews by gathering documentation which is not generally available on the working practices of these organizations." (Emphasis added.) p. 290: "Even if the results from this research cannot be regarded as generally applicable or *generalizable*, some conclusions can be drawn from this multiple case study, contributing to the achievement of a deeper understanding of some internal audit issues.

Gendron and Bedard (2006)	AOS	A series of interviews in 2000 and 2001 conducted in three large Canadian public corporations listed on the Toronto Stock Exchange. A total of 22 individuals were interviewed (7 in corporation A; 8 in corporation B; 7 in corporation C). Another series of interviews were later undertaken in 2004. In footnote no. 7, it is stated that two types of internal documents were also examined at the time of the first-round interview. • In accordance with the views of a number of writers in the domain of qualitative research (e.g., Glaser and Strauss, 1967), the investigation was started with a broad objective in mind. Later, "[s]everal additional themes were incorporated in the interview instrument as the focus of our research was emerging." • Semi-structured interviews were conducted to allow interviewees to express themselves according to their own systems of meaning (Rubin and Rubin, 1995). In the next few sentences after this one, there is a description of the various questions raised. • Dates of interviews and particulars of all interviewees involved in the 2000 and 2001 interviews were laid out in a table. Among the particulars were current position, academic background and professional qualification. • Certain details of 2000/2001 interviews were also provided. Thus, it is mentioned that the length of interviews varied between 45 and 75 minutes. The rest (p. 216): "All interviews were tape-recorded and transcribed, and were attended by either one of the authors. Most of the interviews were face-to-face meetings – apart from two cases for which phone interviews were carried out. • To overcome interviewees' potential reluctance to freely provide information, several measures were incorporated in the interviews (216): "We asked the interviewee for permission to tape the interview, while emphasizing that complete anonymity would be provided to her/him and to the organization, and that no other organization member would examine the interview transcript. Also, participants were told that they would have the opportunity to subsequently verify the accuracy of the transcript and add changes that they feel might be needed to make them comfortable with what they said during the interview." Footnote no. 8 that is attached to this last line has the following to say: "Only minor modifications resulted from the interviewees' revision of their manuscripts."

		Qualitative procedures suggested by Miles and Huberman (1994) were the approach taken in analysing the interview transcripts. It is stated (p. 217): "We used a coding scheme that we developed while reading the transcripts to enhance data sensitivity. Afterwards, for each corporation, a conceptual matrix was prepared to summarize the main themes discussed by interviewees." Right after this sentence, the following was mentioned (p. 217): "We subsequently read literature on actor reflectivity and decided to rely especially on Schutz's analytical concepts to re-examine our interview material in gaining a better understanding of meanings surrounding AC effectiveness." • Regarding the interviews conducted in 2004, the interviewees were composed of three individuals: chairpersons of the audit committees of corporations A, B and C. (In footnote no. 10, it is mentioned too that the individuals chairing the audit committees of A and C in 2004 were also the chairpersons for audit committees in 2000/2001. But for corporation B, the individual chairing the audit committee was not a member of the audit committee during 2000/2001.) Thus, it is stated (pp. 216-7): "Although our second-round interviews are limited and possibly biased towards chairpersons' viewpoints …" • For the analysis of the 2004 interviews, the same type of qualitative procedures used for 2000/2001 interviews were applied.	
Gendron, Cooper and Townley (2007)	AOS	140 semi-structured interviews conducted between 1994 and 2000 with Alberta public servants. • There is a description of the different types of civil servants interviewed. In a footnote no. 4, there is a description too of the questions asked. • The length of time taken up for interview was mentioned with an additional remark that "… almost all were recorded and transcribed." • The analysis of interview data was 'supplemented' by two interviews – each lasted about three hours – with four auditors from the Office of the Auditor General of Alberta. On these interviews, the following was mentioned (p. 4): "These were loosely structured and centred on the history of efficiency auditing in the Office." • Some other interviews were also undertaken (to help in the analysis?) (p. 4): "We also interviewed three individuals involved in the Alberta Financial Review Commission, and two program evaluators working in the Albertan public sector."	footnote 27 which is the last one in the paper and which is attached to a paragraph second last (under the conclusion section): "When Office members read an earlier draft of this paper, they were most upset about the suggestion that they were involved in policy-making. Their concern reflects a popular view that scientists (and, we would add, auditors) are objective whereas policy-making is subject to politics biased by interests. The strained relationship between the Office and us impacted the development of the paper.

		• Other research methods applied? On page 4, it is stated: "Finally, we examined documents published by the Office and Treasury over the period of the research, as well as observing annual, half day, presentations between 1992 and 2005 by senior members of the Office to senior public servants about the role of the Office and its views on accountability, audit and performance measurement.	Although we felt a need for a more comprehensive investigation of the production and *validation* of inscriptions within the Office, it is unlikely that we would have been allowed to carry out any additional interview with Office members. Researchers with better relationships with government auditors may wish to extend our work in this respect."
Kwok and Sharp (2005)	AAAJ	Qualitative or interpretive approach with two modes of data collection: in-depth interviews and document review (archival analysis). • The two modes of data collection (p. 80) "… provides for triangulations to shed more insights into a complicated process." • A delineation of the overall or grand tour question (Werner and Schoepfle, 1987) guiding the study. There is next the listing of 11 sub-questions (p. 80) "… as recommended by qualitative researchers (e.g. Miles and Huberman, 1984)." These questions were the ones raised during interviews taking place over a period of seven months from February to August 1998. • Each interview lasted between one to two hours for a total of 30 interviews. • There is a listing of two criteria in the selection of a group of interviewees for in-depth interviews. For other interviewees, (p. 81) "… other people who are also information-rich with regard to the research goals were selected …" Also, (p. 81) "… new leads were added during field visits, taking advantage of opportunities to talk to other relevant subjects." A footnote attached to this very sentence says: "These strategies or criteria correspond to what Patton (1990) called criterion sampling strategy, snowball or chain sampling, and opportunistic strategy. The paper's appendix list interviewees by board seat and other capacities. • The primary criterion for terminating the interviews was data saturation or redundancy (Lincoln and Guba, 1985) (p. 81) "… when an additional interview did not yield any significant new insights." • Approach during interviews: King's (1994) qualitative research interview. There is a description of what it entails. • Computer software NUD*IST was used (p. 81) "[t]o systemize the analysis of the transcripts."	

		• For document review, the gathering of the four key documents began when data collection started in February 1998. Coding process was explained. It is also mentioned early on that (p. 81) "…the document review of the comment letters used content analysis, which seeks to place narrative text into categories in order to derive conclusions about thematic content." Right after this very sentence, this is mentioned (p. 81): "As stated by Kassarjian (1977), content analysis is no better than its categories (i.e. classification scheme) and that such a scheme must reflect the purpose of the study." Elsewhere it is mentioned that the elementary form of content analysis applied does not measure the relative importance of each category. This is because (p. 82): "Each category is only counted once per letter, regardless of how many times the respondent mentioned it." • Finally, it is mentioned that literature review leads to three specific categories of interest. Next, computer key word searches of document of interest were done using certain key words that fall under these categories. Following this computer searching, the output was compared with manual searching. On page 82: "As a computer search is essentially mechanical in nature and does not replace a researcher's interpretative scrutiny, the coding was compared with the researchers' own detailed reading and evaluation of each comment letter."
Unerman (2003)	AF	Qualitative content analysis is used to collect and summarize the empirical data. *Note*: Though predominantly qualitative content analysis in approach, on page 428, it is stated that "[t]he volume of each individual disclosure (in terms of proportion of a page) has also been recorded and used in the analysis when considered relevant." • To conduct the content analysis, all documents published at UK parent company level as part of Shell's annual reporting cycle between 1950 and 1965 were read at the Shell Archive in London. • On a research instrument, various items were recorded including the type of document in which disclosure had been published, the year to which the document is published, the category of disclosure (which comes in five types) and a summary of the message conveyed by the disclosure. • Next, data from research instruments were entered onto a computerised database. The manner the database was prepared is mentioned in the following manner (p. 429): "… with one record per disclosure, which facilitated flexibility in sorting and analysing the data."

		• From the database, a synopsis of relevant disclosures was compiled. This synopsis was next analysed against key events occurring between 1950 and 1965. This analysis in turn was informed by classical political economy of accounting theory (PET). So, in this study, the form of content analysis followed methods used in previous empirically based classical PET studies of narrative disclosures in company accounts (Adams and Harte, 1998; Neimark, 1992), but was extended to cover a wider range of reports than just the annual reports (Unerman, 2000).
Siti-Nabiha and Scapens (2005)	AAAJ	Longitudinal, interpretative case study of a gas processing company located in an East Asian country. • The research involves 13 visits to the organisation over a period of five years. The initial phase was conducted over a six-month period from July to December 1988 (seven visits). Follow-up visits were made in 2000 (two visits) and also from late 2001 to early 2003 (four visits). The research visits ranged from one to five days in the organisation. • Forty-eight interviews were conducted with 37 organisational members coming from three levels of the organisation. The paper provides an appendix listing the job titles of those interviewed. In the note to the appendix, it is mentioned that there were changes in the positions among those interviewed in the later phases of the research. It also says this: "However, their new positions are not stated, as this is not of particular relevance to the case findings." • Out of 37 organisational members interviewed, eight were interviewed more than once. It is also noted that "[m]ost of the interviews lasted between 1.5 and 2 hours." Also, most of the interviews took place in an important section (a 'key division') which "… operates almost as self-contained unit …" of one of the three levels of the organisation. So, the paper is related primarily to that section which (p. 49) "... has the largest number of staff and resistance to KPIs was apparently greatest …" With these details mentioned, next this is stated (p. 49): "… the persons quoted in this paper were from the plant division, unless otherwise stated." • Evidence from interviews was reinforced by documentary evidence, observations and informal conversations both inside and outside organisation. Various company briefings and training workshops were also attended.

APPENDIX B
Baxter and Eyles's (1997) Table II

Criteria	Definition	Strategies / Practices to Satisfy Criteria
Credibility	Authentic representations of experience	Purposeful sampling
		Disciplined subjectivity/ bracketing
		Prolonged engagement
		Persistent observation
		Triangulation
		Peer Debriefing
		Negative case analysis
		Referential adequacy
		Member checking
Transferability	Fit within contexts outside the study situation	Purposeful sampling
		Thick description
Dependability	Minimization of idiosyncrasies in interpretation	Low-inference descriptors,
		mechanically recorded data
	Variability tracked to identifiable sources	Multiple researchers
		Participant researchers
		Peer examination
		Triangulation, inquiry audit
Confirmability	Extent to which biases, motivations, interests or perspectives of the inquirer influence interpretations	Audit trail products
		Thick description of the audit process
		Autobiography
		Journal/notebook

Shenton's (2004) Chart

Quality Criterion	Possible Provision Made by Researcher
Credibility	Adoption of appropriate, well recognised research methods
	Development of early familiarity with culture of participating organi-
	Random sampling of individuals serving as informants
	Triangulation via use of different methods, different types of infor-
	different sites
	Tactics to help ensure honesty in informants
	Iterative questioning in data collection dialogues
	Negative case analysis
	Debriefing sessions between researcher and superiors
	Peer scrutiny of project
	Use of "reflective commentary"
	Description of background, qualifications and experience of the re-
	Member checks of data collected and interpretations/theories formed
	Thick description of phenomenon under scrutiny
	Examination of previous research to frame findings
Transferability	Provision of background data to establish context of study and de-
	description of phenomenon in question to allow comparisons to be
Dependability	Employment of "overlapping methods"
	In-depth methodological description to allow study to be repeated
Confirmability	Triangulation to reduce effect of investigator bias
	Admission of researcher's beliefs and assumptions
	Recognition of shortcomings in study's methods and their potential
	In-depth methodological description to allow integrity of research
	scrutinised
	Use of diagrams to demonstrate "audit trail"

APPENDIX C
Baxter and Eyles's (1997) List of Eight Questions

- What was the natural history of the research?
- What data were collected and by what methods?
- How was the sampling done?
- How was the data analysis done?
- What results are presented?
- How credible and dependable are the data-construct links?
- How credible is the theory/hypothesis?
- How transferable are the findings?

May and Pope's (2000) List of Questions

- Worth or relevance Was this piece of work worth doing at all? Has it contributed usefully to knowledge?

- Clarity of research question If not at the outset of the study, by the end of the research process was the research question clear? Was the researcher able to set aside his or her research preconceptions?

- Appropriateness of the design to the question Would a different method have been more appropriate? For example, if a causal hypothesis was being tested, was a qualitative approach really appropriate?

- Context Is the context or setting adequately described so that the reader could relate the findings to other settings?

- Sampling Did the sample include the full range of possible cases or settings so that conceptual rather than statistical generalisations could be made (that is, more than convenience sampling)? If appropriate, were efforts made to obtain data that might contradict or modify the analysis by extending the sample (for example, to a different type of area)?

- Data collection and analysis Were the data collection and analysis procedures systematic? Was an "audit trail" provided such that someone else could repeat each stage, including the analysis? How well did the analysis succeed in incorporating all the observations? To what extent did the analysis develop concepts and categories capable of explaining key processes or respondents' accounts or observations? Was it possible to follow the iteration between data and the explanations for the data (theory)? Did the researcher search for disconfirming cases?

- Reflexivity of the account Did the researcher self consciously assess the likely impact of the methods used on the data obtained? Were sufficient data included in the reports of the study to provide sufficient evidence for readers to assess whether analytical criteria had been met?