International Journal of Academic Research in Accounting, Finance and Management Sciences
Vol. 7, No.3, July 2017, pp. 157–170
E-ISSN: 2225-8329, P-ISSN: 2308-0337

www.hrmars.com

On Ensuring Rigour in Accounting Research

Adzor IBIAMKE[1]
Clement C. M. AJEKWE[2]

[1,2]*Department of Accounting, Faculty of management sciences, Benue State University, Makurdi, Nigeria,*
[2]*E-mail: cajekwe@bsum.edu.ng*

Abstract *Researchers in accounting face a challenge of ensuring that their research findings are rigorous, relevant and trustworthy. In the absence of these qualities, academic research findings are irrelevant or unacceptable. This paper discusses what constitutes "rigour' and outlines the methods of ensuring it in quantitative, qualitative and mixed studies in the social sciences generally and particularly accounting. The paper being wholly conceptual utilises literature review to assess criteria to evaluate rigour under the commonly used research approaches. The paper provides an in-depth discussion on validity and reliability issues with their threats under quantitative, qualitative and mixed research approaches. The study concludes with observations that rigour in mixed research is not a mere summation of rigour in quantitative and qualitative components of the mixed design. The paper also concludes that research rigour is enhanced through validity and reliability addressed differently with alternative terms under the three different research approaches in accounting.*

Key words Accounting research, reliability, rigour, validity, quantitative, qualitative, mixed research

DOI: 10.6007/IJARAFMS/v7-i3/3284 URL: http://dx.doi.org/10.6007/IJARAFMS/v7-i3/3284

1. Introduction

The extent to which results of research findings contribute to scientific knowledge depends on their quality evaluated in terms of relevance and rigour (Nørreklit, 2014). Relevance relates to how science and society benefits from the research today and in the future. This suggests that research should have a potential societal impact or relevance (Bouter, 2008). As suggested in the literature, the extent to which a research is judged to be relevant depends on the general assessment criteria of rigour or trustworthiness. Rigour is the ability to reason in a stringent logical manner. It ensures that research findings are intellectually accurate, thorough, and credible; without rigour, research findings are meaningless (Morse *et al*., 2002). Thus, researchers employing quantitative, qualitative or mixed methods acknowledge the importance of rigour (Heale and Twycross, 2015). In quantitative methods, rigour is evaluated with respect to reliability and validity (Roux, 2016). Reliability refers to the extent to which studies can be replicated. It requires that a researcher using the same methods can obtain consistent or equivalent results as those of a prior study. Validity on the other hand, relates to the extent to which measures reflect the phenomena being studied. It pertains to whether a researcher is measuring the right phenomenon; and measuring it holistically.

Unlike quantitative research, rigour in qualitative research is appraised in terms of 'trustworthiness' whose criteria have been a matter of continuous debate and discussion (e.g., Denzin and Lincoln, 2005; Boeije, 2010). The debates and discussions arise because of the different philosophical assumptions of quantitative and qualitative research methods. In spite of the contention regarding the trustworthiness criteria, dependability, credibility, transferability, confirmability, and auditability are widely applied as the standard to assess rigour in qualitative research. Integrating quantitative and qualitative methods (i.e., mixed methods) in the same study further complicates the debates and discussions of rigour.

The clear challenge of researchers is to identify the most important criteria for establishing research rigour ('trustworthiness') in quantitative, qualitative and mixed research approaches with specific reference to accounting research. This conceptual paper discusses what constitutes "rigour" and assembles the

criteria for assessing it under the three different approaches for researchers in social sciences; particularly in accounting.

2. Approaches to accounting research

There are three approaches (qualitative, quantitative and mixed) to research in the social sciences including accounting. These approaches are neither distinct categories, polar opposites, nor dichotomies (Cresswell, 2014). Instead, they represent different ends on a continuum (Newman and Benz, 1998). The frequently mentioned distinction between qualitative and quantitative research is framed in terms of using narratives (qualitative) rather than numbers (quantitative), or using closed-ended questions (quantitative) rather than open-ended questions (qualitative).

Quantitative studies measure and establish numbers, quantity, amounts, frequencies, and intensity; it is based on positivism, a philosophical view that there is a single truth or reality (Higgins & Green, 2008). It is a formal, objective, deductive approach which focuses on the measurement of relationships between variables with a view to building models that can predict outcomes. Quantitative researchers guard against bias; they control for alternative explanations, and are therefore able to generalize the findings. Quantitative research tests hypotheses and theories by examining the relationships among measurable variables. The final written research report has a set structure consisting of introduction, literature and theory, methods, results, discussion and a summary/conclusion.

Qualitative research explores and understands the meanings individuals or groups ascribe to a social or human problem. Qualitative studies collect, analyse and interpret narratives and data that are not easily reduced to numbers. It is based on interpretivism – a philosophical view that there are multiple truths or realities (Higgins & Green, 2008). Thus, knowledge depends on the perspective of the researcher and its existence is understood by collecting words (Denzin and Lincoln, 2000). The final written report has a flexible structure. Those who engage in this form of inquiry support a way of looking at research inductively; focusing on individuals' viewpoints in their environments, circumstances and situations (Cresswell, 2014).

Mixed methods research is an approach to inquiry that integrates both quantitative and qualitative data in one study. The core assumption of mixed methods research is that the combination of both approaches provides a more complete understanding of a research problem than either approach alone. To understand these worldviews (approaches) more clearly, consider the "opposing" philosophical assumptions as presented in Table 1.

Table 1. A comparison of three approaches to social science research

Assumptions	Quantitative (Positivism)	Qualitative (Naturalism)	Mixed
Ontology (*nature of reality*)	A single reality exist beyond ourselves 'out there'	Multiple realities are constructed through our experiences and interactions with others	Reality exists between the participants, the researcher and communities/ individuals being studied. Often a subjective-objective reality emerges.
Epistemology (*how knowledge is acquired*)	Realities can only be approximated. But it is constructed through research and statistics. Interaction with research subjects is kept to a minimum. Validity comes from peers not participants.	Reality is co-constructed between the researcher and the researched and shaped by individual experiences.	Co-created findings with multiple ways of knowing.
Axiological beliefs *(role of values*)	Researchers are independent of what is being studied – hence there is no question about axiology. Researcher's biases need not be expressed in a study	Researcher is immersed in the study; his/her values are made explicit, influencing the research or measures employed in one way or the other.	Respect for indigenous values; values need to be problematised and interrogated

Assumptions	Quantitative (Positivism)	Qualitative (Naturalism)	Mixed
Methodological assumptions (*approach to inquiry*)	Use of scientific method and writing. New knowledge is created through deductive methods such as testing of theories, specifying important variables, making comparisons among groups.	More of a literary style of writing. Use of an inductive method of emergent ideas through consensus obtained through methods such as interviewing, observing, and analysis of texts.	Use of collaborative processes of research; political participation encouraged; questioning of methods; highlighting issues and concerns.
Methods	Predetermined, Closed-ended questions, Performance, attitude, observation and census data, Statistical analysis	Emerging methods, open-ended questions, open-ended interview and audiovisual data, field observation, document data, text and image analysis	Both predetermined and emerging methods, both open- and closed -ended questions, multiple forms of data drawing on all possibilities, Statistical and text analysis
Rhetoric (*language and the writing approach of the researcher*)	Impersonal and expressive through statistical results or quantified descriptors – internal validity, reliability, generalisability and objectivity	Personal, literary, and based on definitions that evolve during a study. Use of metaphors, and sometimes personal reference using the first-person pronoun, "I," and storytelling. Focus is on credibility, transferability, dependability and confirmability	Combination of personal and impersonal, formal and informal, subjectivity, discourses, discourse analysis, reflexivity, subject and self, and deconstruction language depending on the nature of data at a given stage.
Use these practices of research, as the researcher	Tests or verifies theories or explanations; identifies variables to study; relates variables in questions or hypotheses; uses standards of validity and reliability; observes and measures information numerically; uses unbiased approaches; employs statistical procedures	Positions self, collects participants' meanings; focuses on a single concept or phenomenon; brings personal values into the study; studies the context or setting of participants; validates the accuracy of findings; makes interpretations of the data; and creates an agenda for change/ reform	Collects both quantitative and qualitative data; develops a rationale for mixing; presents visual picture of the procedure in the study; employs the practices of both qualitative and quantitative research

Source: Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed). London: Sage Publications.

It is apparent that qualitative and quantitative research operates in different domains, with different missions and agendas. Rigour is demonstrated in different ways and by different terms depending on the research approach. Research rigour is enhanced and addressed differently with alternative terms under the three different research approaches as summarised in Table 2.

Table 2. Framework for rigour in the three research approaches

Quantitative	Mixed	Qualitative
Design-Related Elements		
Validity	Design suitability, design adequacy/fidelity, within design consistency, analytic adequacy	Credibility, trustworthiness, plausibility, authenticity, applicability
Measurement-Related Elements		
Reliability	Sample integration, Inside-outside, weakness minimization, Sequential, Conversion, Paradigmatic mixing, Commensurability, Multiple validities, and Political	Consistency Triangulation Auditability Credibility Confirmability

Quantitative	Mixed	Qualitative
Inference-Related Elements		
Statistical conclusion Validity	Interpretive consistency Theoretical consistency Interpretive agreement Interpretive distinctiveness Integrative efficacy	Giving voice, peer debriefing, triangulation, reflexive journaling, persistent observation, dependability audit, articulating decisions, member checking

Source: Leech, N., Dellinger, A., Brannagan, K. & Tanaka, H. (2010). Evaluating mixed research studies: A mixed methods approach. *Journal of Mixed Methods Research,* 4(1), 17-31.

3. Rigour in quantitative accounting research

In this section, rigour will be discussed under two sub-headings: validity and reliability.

3.1. Validity in Quantitative Research

Validity means the extent to which an instrument measures *what* it is supposed to measure and *how well* it does so (Smith, 1991). The concern for validity in quantitative studies is the avoidance of type I and type II errors (Long and Johnson, 2000). A type I error is the error of rejecting a true null hypothesis (a 'false positive'); while a type II error is the error of incorrectly retaining a false null hypothesis (a 'false negative'). In legal terms, type I and type II errors are equivalent to declaring a 'guilty person innocent' and an 'innocent person guilty' respectively. Validity is broadly grouped into internal or external validity.

3.1.1 Internal validity

Internal validity is the extent to which changes in the dependent variable results from changes in the independent variable(s) (Abernethy *et al.*, 1999). It is a duty of a researcher to adequately control the research process in such a manner that the resultant effect on the dependent variable is wholly and exclusively caused by the independent variables. In experimental research, an instrument should control for the possible effect of extraneous factors. The less there is any chance of extraneous (confounding) factors' effect in a study, the higher is internal validity achieved. Thus, in accounting research the most important question is whether valid conclusions can be drawn using a given research design and controls (Ryan *et al.*, 2002). An internally valid study is that in which conclusions are drawn from a set of observations with little ambiguity. Four types of internal validity are content validity, criterion-related validity, and construct validity (Eby 1993, Punch 1998). Cook and Campbell (1979) include statistical conclusion validity as a precursor to the entire internal validity.

3.1.1.1 Content validity

Content validity refers to the extent to which a measure accurately covers all the domains of interest. Content validity measures whether the contents included on the scale are 'appropriate' and 'thorough' enough to adequately represent the concept of interest. Content validity requires that the concept domain is made clear and the measure(s) fully represent the domain (Bollen, 1989). Content validity is often judged by logical, rather than statistical evidence. Subject matter experts are asked to determine whether the substance of the instrument is reasonably related to and covers all the domains of the concept. In accounting research, the problem with content validity is that, most concepts are not directly observable; neither do they have a consensus definition (Lukka and Vinnari, 2014). This makes the content domain ambiguous. Consequently, the burden is on the researcher not only to provide a theoretical definition of the concept that is generally accepted (by peers) but also to select indicators that thoroughly cover the domain and dimensions of the concept.

3.1.1.2 Criterion-related validity

Criterion-related validity means that a particular measure corresponds with its referent (criterion). It shows the correlation between a measure and some criterion variable of interest. Literally, it is the comparison of a measure against a single measure that is supposed to be a direct measure of the concept

under study. If there is high correlation, the measure is valid for the criterion; if otherwise, it is not valid for the criterion and thus not useful for the particular purpose. Criterion-related validity can take any of the two forms: concurrent and predictive validity. When the criterion exists at the same time as the measure, we talk about *concurrent validity*. When the criterion occurs in the future, we talk about *predictive validity*. For instance, an aptitude test scores to predict employee's future success on the job is a predictive validity. Criterion validity is measured in three ways: (a) convergent validity—an instrument is highly correlated with instruments measuring similar variables; (b) divergent validity—an instrument is poorly correlated with instruments that measures a similar thing and (c) predictive validity—instrument have high correlations with future criterion. A measure is valid if by these tests, the measure is positively correlated with the other measures. If a negative correlation is found, the test offers existence of discrimination (McDonald, 2005).

3.1.1.3 Construct validity

Construct validity involves demonstrating relationships between the study constructs and the theory expectations about those constructs. It is the congruence between the study's result and the theoretical underpinnings guiding the particular research. Quantitative studies are driven by an underlying theory; so, construct validity measures the degree to which a test measure performs in accordance with theoretical expectations. Using theory, the researcher formulates theoretical predictions about the existence, direction, and extent of relations among empirical indicators. It requires a construct fitting with a theory; agreement between a study finding and theoretical expectation of the theory guiding it. Construct validation has three steps: (i) a theoretical relationship between the research concepts; (ii) empirical relationships between the measures of the concepts; and (iii) interpretation of empirical evidence. Theory prediction is fundamental in construct validity; consequently, empirical results should be consistent with theoretical expectation. Where empirical results are inconsistent with theoretical predictions then either; (i) the study was based on an inappropriate theory, (ii) the research methods or procedure adopted may have been faulty and (iii) some variables of the study lacked construct validity or reliability (Cronbach and Meehl, 1955; Carmines and Woods, 2005). Construct validity is usually tested using correlation factor analysis, ANOVA multi-trait/multi-method.

3.1.1.4 Statistical conclusion validity

Statistical conclusions validity is the degree in which conclusions reached about relationships from the data are reasonable, credible or believable. It is about the proper use of statistical procedures in analyzing data. Typical threats to statistical conclusion validity are violating the assumptions of statistical procedures such as the independence of observations, low statistical power, and increasing the chance of a type I error by data fishing (Reidardt, 2005). The validity addresses appropriateness of the statistical procedures employed and the degree to which statistical assumptions have been satisfied. It is generally increased with increasing statistical power. This can be achieved by using a larger sample size, raising the alpha level (this can however increase the probability of making Type II error) or increasing the salience of the relationship itself. Furthermore, conclusion validity could be improved through better construction of measurement instruments, increasing the number of questions on a scale or by reducing situational distractions in the measurement context (Trochim, 2006). A lack of statistical validity can as well invalidate external validity because critical assumptions about the model are violated or there is no predictive power about the study model.

Internal validity can be threatened at any stage of a research process – during research design, data collection, analysis and/or interpretation. A good research must therefore take adequate measures in the entire research process to ensure high internal validity. A list and explanation of threats to internal validity and ways of demonstrating construct validity are presented in Table 3.

3.1.2 External validity

External validity means the extent to which results of a study can be generalised to other contexts, situations or people. External validity ensures that the conditions under which the study is carried out are representative of the situations and time to which the results are to apply (Black 1999; Ryan *et al*., 2002). The sample of participants drawn from the population of interest must be representative of that population

at the time of the study. Finally, representative samples should be drawn with reference to relevant variables in the study, such as industry sector and firm size. For this reason, external validity is alternatively called generalised or transferability validity.

Table 3. A list and explanations of types of threats to internal validity

Types of threats	Explanation
Insufficient knowledge	Insufficient knowledge of or logical incongruence between research question, theory, statistical tests and analysis.
History Effects	A history effect occurs when some change other than the experimental treatment occurs during the course of an experiment that affects the dependent variable. History effects are particularly prevalent in repeated measures experiments that take place over an extended time. A special case of the history effect is the *cohort effect*, where a change in the dependent variable occurs because members of one experimental group experienced different historical situations than members of other experimental groups.
Maturation	Changes in the participants as a result of the passage of time or other naturally occurring events like growth and experience leads to maturation threats. Studies taking place over longer time spans may see lower internal validity as subjects simply grow older or more experienced.
Testing Effect	The effect of subsequent tests on the current tests scores is referred to as testing effect (also called *pre-testing effects* because the initial test affects the response of subject in the subsequent experiments or tests). For example, the common belief is that repeated practice with GMAT exam leads to better test taking skills suggesting that a second administration of the GMAT would lead to higher scores, independent of any gain in actual knowledge. Testing effects occur mainly in a before-and-after study or repeated designs.
Instrumentation	Instrumentation refers to the change in calibration of an instrument over time. A change in the wording of questions, a change in interviewers, or a change in other procedures used to measure the dependent variable causes an *instrumentation effect*, which can jeopardize internal validity.
Selection	Selection refers to bias that result from differential selection of respondents for experimental groups. Researchers must make sure the characteristics of the research subjects accurately reflect the population. Furthermore, the key characteristics of the subjects must be distributed in such a way as to create equal groups.
Mortality/Attrition	Mortality/Attrition is the differential loss of respondents from the comparison groups. Attrition occurs when some subjects withdraw from the experiment before it is completed, while mortality occurs where subjects drop from an experimental treatment group disproportionately than from the other groups. Mortality or attrition occurs in studies that take a longer period.
Statistical regression	Statistical regression is the tendency for an unusually high or low score to regress or return to a more usual or mean level on subsequent measures. Statistical regression, occurs when respondents or subjects are selected on the basis of their extreme scores; if the measuring device is not totally reliable.

Source: Researchers' summary

Generalisation is the process of using particular data to infer a general statement that has applicability to other people, settings, or times (Ferguson, 2004). External validity is threatened by the factors presented in Table 4.

Table 4. A list and explanations of types of threats to external validity

Types of threats	Explanation
Population	Population refers to whether inferences can be drawn from a study of a given population. External validity is threatened, where biases or other limitations exist in the accessible population. This can likely happen where the sample size is inadequate and/or not randomly selected.

Types of threats	Explanation
Time	Time is the degree to which the results of a particular study at a point in time can be generalized to other time periods. Where changes in the relationships between variables occur from period to period, the time validity of such a study is said to be low.
Environmental	Environmental validity indicates whether results can be generalized across different settings (places).

Source: Researchers' summary

3.2 Reliability in Quantitative Research

The concept of reliability is defined as: "the extent to which repetition of the study would result in the same data and conclusions"(Goode and Hatt, 1952); "the accuracy or precision of a measuring instrument" (Kerlinger, 1964); "the ability of the instrument to measure consistently the phenomenon it is designed to measure" (Black & Champion, 1976).

A synthesis of these definitions suggests that "reliability" means "repeatability" or "consistency". A measure is reliable if it gives the same result over and over again (provided what is being measured isn't changing). This implies that with the reliability of a research instrument or method, consistent and accurate data or results are produced even if used by different people or at different times. If measurement results are not reliable, it becomes more difficult and precarious to test hypotheses or to make inferences about the relations between variables (Kerlinger, 1964). There are two types of reliability-internal and external.

3.2.1 Internal reliability

Internal reliability refers to the consistency of results across items within a test. There are three main concerns for achieving internal reliability: stability, internal consistency and equivalence. *Stability* is the ability of a measure to remain the same over time despite changing or uncontrollable testing conditions or the state of the respondents themselves. When multiple measurements are taken, reliable measures will all be consistent in their values (Hair *et al.*, 2006).

Internal consistency relates to how well a set of items measure a particular phenomenon within the study. That is, it measures how the items "hang" together as a set, and are capable of independently measuring the same concept so that the respondents attach the same overall meaning to each of the items. For a test to be internally consistent, estimates of reliability are assessed on the average inter-correlations among all the single items within a test. This can be seen by examining if the items and the subsets of items in the measuring instrument are correlated highly.

Equivalence (lack of bias) is to the quality or state of having the similar value. To ensure equivalence, a set of questions that measure the same constructs, knowledge or skill are randomly split into half in several ways, e.g., first half and second half, or by odd and even numbers. If the two halves of the test provide similar results, this would suggest that the test has internal reliability. This approach (also known as the split-half test) is a quick and easy way to establish reliability, especially with large questionnaires in which all questions measure the same construct.

3.2.2 External reliability

External reliability refers to the extent to which questionnaires and psychometric measures vary from one user to another. External reliability is gauged by the test-retest methods (i) to assess the stability of a measure over time and (ii) by different researchers. In the stability test-retest method, the results of participants on two separate occasions are assessed; if the same or similar results are obtained, then, external reliability is established. The timing of the test is important; if the duration is too brief then participants may recall information from the first test which could bias the results. Alternatively, if the duration is too long it is possible that the participants could have changed in some important ways which could also bias the results. In the inter-rater tests, assessments of the degree to which different researchers give consistent estimates of the same behaviour are confirmations of external consistency. Where observer scores do not significantly correlate, then reliability can be improved by: (a) training observers in the observation techniques being used and making sure everyone agrees with them; and (b) ensuring behaviour categories have been objectively defined.

Reliability is measured by the proportion of variability in the true score (rather than some kind of error); it can be examined through the Cronbach's alpha coefficient. A cronbach's alpha coefficient value of 0.9 means 90 per cent of the variability in the observed score is true and 10 per cent is due to error. A reliability of 80 to 90 per cent is recommended. A lower Cronbach alpha means that either the test is too short or the items have very little in common.

In summary, reliability is about consistency of measurement (Bollen, 1989), or stability of measurement over a variety of conditions in which basically the same results should be obtained (Nunnally, 1978). It is important to note that reliability is a necessary but not sufficient condition of the test of goodness of a measure. For example, one could very reliably measure a concept establishing high stability and consistency, but it may not be the concept that one had set out to measure.

4. Rigour in qualitative research

The terms *reliability* and *validity*, are traditionally used to evaluate rigour in quantitative research, yet they cannot be avoided or compromised in conducting or reporting qualitative research (McKinnon, 1988). However, qualitative researchers (e.g., Lincoln and Guba, 1985; Guba and Lincoln, 1989; Sandelowski, 1986; Tracy, 2010) evaluate rigour in terms somewhat similar but not necessarily identical to the concepts of validity and reliability.

4.1. Validity in Qualitative Research

Discussions regarding rigour in qualitative research are on-going; Lincoln and Guba's (1985) criteria (credibility, transferability, dependability and confirmability) are considered the 'gold standard' (Tashakkori and Teddlie, 1998). Credibility, trustworthiness and neutrality parallel internal validity, transferability resembles external validity, dependability parallels reliability and confirmability resembles external validity/objectivity. Concepts relating to validity are discussed in this section; and those corresponding to reliability are discussed in the next section.

4.1.1 Trustworthiness or credibility

Trustworthiness or credibility is the overriding basis for evaluating rigour in qualitative research (Lincoln and Guba, 1985). It is concerned with whether the study actually measures or tests what is intended. To be credible, conscious efforts are made to establish confidence in the accuracy and interpretation of data (Carboni, 1995). Credibility seeks to answer questions such as: Do the results of the research reflect the experience of participants or the context in a believable way (Lincoln and Guba, 1985)? Does the explanation fit the description? (Janesick, 1994) What is the assurance that interpretations from the study are trustworthy and reveal some truth outside the investigator's bias or preferences? (Thorne, 1997)

Credibility of research findings are enhanced by triangulation approaches. *Triangulation* is a validity procedure where researchers seek convergence among multiple and different sources of information to form themes or categories in a study (Creswell and Miller, 2000). Denzin (1978) identified four types of triangulation: across data (data triangulation), sources (triangulation by participants), interview, observations, documents (method triangulation) and among different investigators (investigator/evaluator triangulation, also known as peer debriefing). As a validity procedure, triangulation enhances the credibility of research findings. Overall, validity threats of distortion, bias, and inadequate portrayal of the participants/phenomenon are addressed through credibility, ultimately contributing to quality in qualitative research (Creswell and Miller, 2000).

4.1.2 Transferability

Transferability means the degree to which the results of qualitative research can be generalized or transferred to other contexts or settings. It is established by providing readers with evidence that the research findings could be applicable to other contexts, situations, times, and populations (Lincoln and Guba, 1985). Such evidence is achieved by providing a rich and thick explanation of where the interviews occurred (research sites), how the participants in the study were recruited, when and where the interviews were conducted, and other aspects of data collection that help provide a richer and fuller understanding of

the research setting. Because the findings of the research are not merely reported but described in vivid details, outside researchers and readers are able to make their transferability judgements.

4.1.3 Confirmability

Confirmability connotes the extent to which others can confirm the findings in order to ensure that the results reflect the understandings and experiences from observed participants, rather than the researcher's own preferences. Confirmability is important because qualitative research tends to assume that each researcher brings a unique personal perspective to the study. In addition to triangulation methods, Lincoln and Guba (1985) suggest using an 'inquiry audit' or "audit trail" to enhance confirmability. An audit trail is established by researchers documenting the inquiry process through journals and memos, keeping a research log of all activities, developing a data collection chronology, and recording data analysis procedures clearly. In this way, individuals outside the project (external auditors) can study and review the documentation making the narrative account credible. On the whole, as can be construed from this section, it is of utmost importance that the credibility (validity) of the account be conveyed in a qualitative study.

4.2. Reliability in Qualitative Research

Reliability in quantitative research refers to the consistency, stability, and dependability of a test or testing procedure (Sandelowski, 1986). However, qualitative research seeks to produce credible knowledge of interpretations on organisations, management accounting processes and understandings, with an emphasis more on uniqueness of human situations and variations in an experience rather than identical repetition of tests or testing procedure (Parker, 2012; Wahyuni, 2012). In these circumstances, auditability has been proposed to be the criterion of rigour relating to consistency of qualitative findings (Guba and Lincoln, 1981).

4.2.1 Auditability

Auditability is achieved when the researcher leaves a clear decision trail of the study from its beginning to the end. Sandelowski (1986) specified achievement of auditability by describing, explaining, or justifying: (i) how the researcher became interested in the subject matter of the study; (ii) how the researcher views the subject of study; (iii) the specific purpose(s) of the study; (iv) how the subjects or pieces of evidence came to be included in the study and how the participants were recruited; (v) the reciprocal impact between the subjects or evidence and the researcher; (vi) how the data were collected; (vii) how long data collection lasted; (viii) the nature of the settings in which data were collected; (ix) how the data were reduced or transformed for analysis, interpretation and presentation; (x) how various elements of the data were weighted; (xi) the inclusiveness and exclusiveness of the categories developed to contain the data and the specific techniques to determine the truthfulness and applicability of the data.

4.2.2 Dependability

Dependability corresponds to the notion of internal reliability which promotes consistency, and repeatability or replication Dependability concerns taking into account all the changes that occur in a setting and how these affect the way research is being conducted. Dependability can be achieved by a detailed explanation of the research design and process to enable future researchers to follow a similar research framework. It should be noted that the application of the research model by a future researcher is not necessarily targeted at getting a similar result. Enhancing dependability can be achieved by presenting detailed and step-by-step explanation of the research processes undertaken, as well as providing the main instruments used to gather empirical data, for instance, the list of interview questions. Qualitative research is often criticized as biased, small scale, anecdotal, and/or lacking rigour; however, when it is carried out properly it is unbiased, in depth, valid, reliable, credible and rigorous.

5. Rigour in mixed methods research

Mixed methods integrate quantitative and qualitative methods within a single study (Johnson *et al.*, 2007, Tashakkori and Creswell, 2007). Given the different epistemological and ontological assumptions of

quantitative and qualitative methodologies, assessing rigour in a mixed methods study is complex and requires additional consideration even as it has been suggested that there is poor consensus regarding what constitutes rigour in mixed methods (Brown *et al.*, 2015). It is envisaged however that more detailed description of data collection and analysis, integration, inferences and justification of mixed methods would be required (Seale, 1997; Whittemore *et al.*, 2001). Reasoning that rigour in mixed is the summation of rigour in qualitative and quantitative research is not tenable because a mixed research method is more than the sum of its parts (Curry & Nunez-Smith, 2015).

A number of authors, particularly Teddlie and Tashakkori (2009), Tashakkori and Teddlie (2008) and Onwuegbuzie and Johnson (2006) have developed integrative and legitimation frameworks that enable assessment of rigour in mixed methods studies. The integrative framework by Teddlie and Tashakkori (2008) and Tashakkori and Teddlie (2009) as summarised by Ihantola and Kihn (2011) as: (i) *Design quality* - the design suitability, design adequacy or fidelity, analytic adequacy and consistency of procedures within the design. "Design suitability" has to do with justification of the appropriateness of a mixed design. "Design adequacy"/fidelity is concerned with whether the components of the design are implemented adequately. "Analytic adequacy" addresses the adequacy and appropriateness of the techniques of data analysis. (ii) *Interpretive rigour* entails: interpretive consistency, theoretical consistency, interpretive agreement, interpretive distinctiveness and integrative efficacy. "*Interpretive consistency*" has to do with the consistency of inferences with each other and with the results of data analysis. "*Theoretical consistency*" addresses whether each inference is consistent with current theories in the academic field and/or with empirical findings of other studies. "*Interpretive agreement*" refers to the consistency of interpretations across scholars and the participants' construction of reality. Threats to interpretive agreement exist if other scholars do not agree that the inferences are the most plausible interpretations of the findings, and the interpretations do not make sense to the participants of the study. "*Interpretive distinctiveness*" is the degree to which the inferences are distinct from other possible interpretations of the results and the rival explanations are eliminated. Interpretive distinctiveness is not demonstrated if there are other plausible explanations for the findings. Finally, "*integrative efficacy*" is the degree to which inferences made in each strand of a mixed methods study are effectively integrated into a theoretically consistent meta-inference.

The four previous criteria related to interpretative rigour are applicable to both qualitative and quantitative parts of the research and to the meta-inferences that emerge when the inferences of the two or more parts are integrated. By contrast, integrative efficacy is unique to meta-inferences in mixed methods. It is concerned with the degree to which a mixed methods researcher adequately integrates the findings, conclusions, and policy recommendations gleaned from each of the two strands and meaningful conclusions can be made of them.

The legitimation framework developed by Onwuegbuzie and Johnson (2006) is a continuous iterative and interactive process that should occur at each stage of the mixed research process. Legitimation means ''making inferences that are credible, trustworthy, dependable, transferable, and/or confirmable'' (Onwuegbuzie and Johnson, 2006). The types of legitimation and examples of their threats are summarised in Table 5.

Table 5. The legitimisation framework

Legitimation Type	Examples of Threats
Sample integration: The extent to which the relationship between the quantitative and qualitative sampling designs yields quality meta-inferences	Mismatch between quantitative and qualitative samples
Inside-outside: The extent to which the researcher faithfully presents and appropriately utilizes the insider's view and the observer's views for purposes such as description and explanation	The imbalance between insider's and outsider's views (e.g. the researcher has failed to maintain a well-informed and balanced perspective when collecting, analysing, and interpreting what the whole set of qualitative and quantitative data mean)
Weakness minimization: The extent to which the weakness from one approach is compensated by the strengths from the other approach	Careless assessment of threats to and weaknesses of quantitative and qualitative parts of research. Deficiencies in compensating the weaknesses by the strengths

Legitimation Type	Examples of Threats
Sequential: The extent to which one has minimized the potential problem wherein the meta-inferences could be affected by reversing the sequence of the quantitative and qualitative phases	The sequencing itself is a threat if the results and interpretations are different if the order of the quantitative and qualitative phases are reversed
Conversion: The extent to which quantizing or qualitizing yields quality meta-inferences	Counting pitfalls associated to verbal counting, misleading, a contextual and over- counting. Over-generalizations and representations of people that are unrealistic
Paradigmatic mixing: The extent to which the researcher's epistemological, ontological, axiological, methodological and rhetorical beliefs that underlie the quantitative and qualitative approaches are successfully (a) combined or (b) blended into a usable package	Competing dualisms of paradigmatic assumptions: the researcher does not make her/his paradigmatic assumptions explicit and does not conduct the research according to the stated assumptions
Commensurability: The extent to which the meta-inferences made reflect a mixed worldview based on the cognitive process of Gestalt switching and integration	Lack of cognitive and empathy training of researchers and their inability to make Gestalt switches
Multiple validities: The extent to which addressing legitimation of the quantitative and qualitative components of the study result from the use of quantitative, qualitative, and mixed validity types, yielding high-quality meta-inferences	Threats to the quality of quantitative and qualitative parts of the study
Political legitimation: The extent to which the consumers of mixed methods research value the meta-inferences stemming from both the quantitative and qualitative components of a study	Value or ideologically-based conflicts when different quantitative and qualitative researchers collaborate in a mixed methods study The contradictions and paradoxes when qualitative and quantitative data are compared and contrasted. The difficulty in persuading consumers of mixed methods research to value the meta-inferences stemming from both the qualitative and quantitative findings

Source: Onwuegbuzie and Johnson (2006). The validity issue in mixed research. *Research in the Schools,* 13(1), 48-63.

6. Conclusions

This paper emphasises that rigour is required in all studies irrespective of the research approach chosen by the researcher. Without rigour, research findings are meaningless. This paper explained how rigour is demonstrated in different ways and in different terms depending on the research approach. Quantitatively, rigour is enhanced using validity and reliability. In the qualitative approach rigour is about trustworthiness of the findings achievable through credibility, transferability, dependability, and confirmability criteria. Finally, rigour in mixed research is not just a mere summation of rigour in quantitative and qualitative components. Both an integrative and legitimisation frameworks of assessing rigours should be applied

References

1. Abernethy, M., Chua, W., Luckett, P., Selto, F. (1999). Research in managerial accounting: Learning from others' experiences. *Accounting and Finance*, 39, 1-27.
2. Black, T. (1999). *Doing Quantitative Research in the Social Sciences*. London: SAGE.
3. Boeije, H. (2010). *Analysis in qualitative research.* London: Sage Publications.
4. Bollen, K. (1989). *Structural equations with latent variables*. London: John Wiley & Sons.
5. Bouter, L. (2008). Knowledge as public property: The societal relevance of scientific research. Available at http://www.oecd.org/site/eduimhe08/41203349.pdf
6. Brown, K., Elliot, S., Leatherdale, S., Robert-Wilson, J. (2015). Searching for rigour in the reporting of mixed methods population health research: A methodological review. *Health Education Research,* 30(6), 811–839.

7. Bryman, A, Becker, S. Sempik, J. (2008). Quality criteria for quantitative, qualitative and mixed methods research: A view from social policy. *International Journal of Social Research Methodology,* 11, 261–76.

8. Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research,* 6**,** 97-113.

9. Bryman, A. (2012). *Social research methods* (4th edn.). Oxford: Oxford University Press.

10. Carboni, J. (1995). A Rogerian process of inquiry. *Nursing Science Quarterly*, 8, 22-37.

11. Carmines, E. & Woods, J. (2005). Validity assessment. In K. Kempf-Leonard (ed) *Encyclopedia of social measurement,* 3, 933-937.

12. Cook, T. and Campbell, D. (1979). *Quasi-experimentation: Design and Analysis Issues for Field Settings.* Boston: Houghton Mifflin.

13. Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104.

14. Creswell, J. and Miller, D. (2000). Determining validity in qualitative inquiry. *Theory into Practice;* 39(3), 124-130.

15. Creswell, J. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed). London: SAGE.

16. Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika,* 16(3): 297-334.

17. Cronbach, L. and Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

18. Curry, L. and Nunez-Smith. (2015). *Mixed methods in health sciences research: A Practical Primer.* London: SAGE.

19. Denzin, N. and Lincoln, Y. (eds) (2000). *Handbook of qualitative research* (2nd ed). London: SAGE.

20. Denzin, N. and Lincoln, Y. (2005). *The sage handbook of qualitative research*, (3rd edn). London: SAGE.

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

22. Eby, M. (1993). Validation: Choosing a test to fit the design. *Nurse Researcher*, 1(2), 27-33.

23. Ferguson, L. (2004). External validity, generalizability, and knowledge utilization. *Journal of Nursing Scholarship*, 36(1), 16-22.

24. Golden-Bibble, K. and Locke, K. (1993). Appealing works: An investigation of how ethnographic texts convince. *Organization Science*, 4, 595-616.

25. Guba, E. and Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park, CA: SAGE.

26. Hair, J., Black, W., Babin, B., Anderson, R. and Tatham, R. (2006). *Multivariate data analysis.* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

27. Heale, R. and Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-based nursing*, 18(3), 66-67.

28. Ihantola, E. and Kihn, L. (2011). Threats to validity and reliability in mixed methods accounting research, *Qualitative Research in Accounting & Management,* 8(1), 39 – 58.

29. Janesick, V. (1994). The dance of qualitative research design: Metaphor, methodolatry and meaning.In N. K. Denzin and Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 209-219). Thousand Oaks, CA: SAGE.

30. Johnson, R., Onwuegbuzie A. and Turner, L. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research*, 1(2), 112-133.

31. Kalof, L., Dan, A. and Dietz, T. (2008). *Essentials of social research* (1st ed). Open University Press.

32. Kerlinger, R. (1964). *Foundations of behavioral research* (2nd ed). New York: Holt, Rinehart and Winston.

33. Kuder, G. and Richardson, M. (1937). The theory of the estimation of test reliability. *Psychometrika,* 2*,* 151–160.

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

35. Lincoln, Y. and Guba, E. (1985) (eds). *Naturalistic Inquiry.* Newbury Park, CA: SAGE.

36. Lukka, K. and Modell, S. (2010). Validation in interpretive management accounting research. *Accounting, Organizations and Society,* 35(4): 462-477.

37. Lukka, K. and Vinnari, E. (2014), Domain theory and method theory in management accounting research. *Accounting, Auditing & Accountability Journal,* 27(8), 1308 - 1338

38. Mcdonald, M. (2005). Validity, data sources. In K. Kempf-Leonard (ed) *Encyclopedia of social measurement, 3, pp.* 939-948.

39. McKinnon, J. (1988). Reliability and validity in field research. *Accounting, Auditing and Accountability Journal,* 1(1), 34-54.

40. Morse, J., Barrett, M., Mayan, M., Olson, K., and Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods,* 1(2), 13-22.

41. Leech, N., Dellinger, A., Brannagan K. and Tanaka, H. (2010). Evaluating mixed research studies: A mixed methods approach. *Journal of Mixed Methods Research,* 4(1), 17-31.

42. *Newman*, I. and *Benz*, C. (*1998*). *Qualitative-quantitative research methodology: Exploring the interactive continuum.* Carbondale: University of Illinois Press.

43. Nørreklit, H. (2014). Quality in qualitative management accounting research. *Qualitative Research in Accounting,* 11(1), 29-39.

44. Nunnally, J. C. (1978). *Psychometric theory*. London: McGraw-Hill.

45. Onwuegbuzie, A. and Johnson, R. (2006). The validity issue in mixed research. *Research in the Schools,* 13(1), 48-63.

46. Onwuegbuzie, A. (2003). Expanding the framework of internal and external validity in quantitative research. *Research in the Schools.* 10(1), 71-90.

47. Parker, L.D. (2012). Qualitative management accounting research: Assessing deliverables and relevance. *Critical Perspectives on Accounting,* 23(1), 54-70.

48. Punch, K. (1998). *Introduction to social research*: *Quantitative and qualitative approaches.* London: SAGE.

49. Reidardt, C. (2005). Treatment effect. In K. Kempf-Leonard (ed) *Encyclopedia of social measurement,* 3, 878-882.

50. Rosenthal, R. and Rosnow, R. (1991). *Essentials of behavioural research: Methods and data analysis.* (2nd ed.). London: McGraw-Hill.

51. Roux, C. (2016). Exploring rigour in auto-ethnographic research. *International Journal of Social Research Methodology*. 1-13. http://dx.doi.org/10.1080/13645579.2016.1140965

52. Ryan, B., Scapens, R. and Theobald, M. (2002). *Research method and methodology in finance and accounting,* (2nd ed.). London: Thomson.

53. Sandelowski, M. (1986). The problem of rigor in qualitative research. *Advances in Nursing Science,* 8(3), 27-37.

54. Seale, C. (1997). Ensuring rigour in qualitative research. *European Journal of Public Health,* 7, 379–84.

55. Smith, J. (1990). Alternative research paradigms and the problem of criteria. In E. G. Guba (ed) *The paradigm dialogue* (pp. 167-187). Newbury Park, CA: SAGE.

56. Tashakkori, A. and Teddlie, C. (2010) (eds). *SAGE handbook of mixed methods in social & behavioural research,* (2nd ed). Thousand Oaks, CA: SAGE.

57. Tashakkori, A. and Creswell, J. (2007) Editorial: The new era of mixed methods. *Journal of Mixed Methods research*, 1(1), 3-7.

58. Tashakkori, A. and Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: SAGE.

59. Tashakkori, A. and Teddlie, C. (2003). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (eds.), *Handbook of mixed methods in social and behavioral research* (pp. 671-702). Thousand Oaks, CA: SAGE.

60. Tashakkori, A. and Teddlie, C. (2006). Validity issues in mixed methods research: Calling for an integrative framework. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

61. Tashakkori, A. and Teddlie, C. (2008). Quality of inferences in mixed methods research. In M. Bergman (Ed.), *Advances in mixed methods research: Theories and applications* (pp. 101-119). London: SAGE.

62. Teddlie, C.and Tashakkori, A. (2009) *Foundations of mixed methods research.* London: SAGE.

63. Thorne, S. (1997). The art (and science) of critiquing qualitative research. In J.M. Morse (Ed.), *Completing a qualitative project: Details and dialogue* (pp. 117-132). Thousand Oaks, CA: SAGE.

64. Tracy, S. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry* 16(10), 836-851.

65. Trochim, W. (2006). Qualitative measures. The research methods knowledge base. Retrieved on 9/4/2017 from https://www.socialresearchmethods.net/kb/qual.php

66. Wahyuni, D. (2012). The research design maze: Understanding paradigms, cases, methods and methodologies*, Journal of applied management accounting research*, 10(1), 69 – 80.

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

68. Williams, P. (2014). The myth of rigorous accounting research. *Accounting Horizons,* 28(4), 869–887.