# The Interpretation and use of Mixed Methods Research Within Programme Evaluation Practice

by

Kyeyune Apolo Peter

*Thesis presented in fulfilment of the requirements for the degree*
*Masters of Philosophy in Social Science Methods at the*
*University of Stellenbosch*

Supervisor: Prof. Johann Mouton

Faculty of Arts

Department of Sociology and Social Anthropology

December 2010

# Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2010

# Abstract

The contemporary evaluation literature advocates for and recommends a pluralistic approach to programme evaluation, with some writers contending that the use of multiple and/or mixed methods for the practice is inevitable. The rationale for such an approach encompasses aspects of both the 'technical' and the 'political' requirements of evaluation practice. A review of evaluation research literature underscores the important role of mixed methods research towards realizing richer evaluation findings, and addressing the pragmatic, democratic and political facets of the evaluation practice. However, it is observed that there is a dearth of literature that focuses on how the use of a mixed methods evaluation approach facilitates the realization of richer conclusions or inferences about programme merit/worth. Thus, the overarching aim of the thesis is to establish how the perception and implementation of mixed methods research among evaluation practitioners influences the nature of inferences they make.

This thesis aims at identifying patterns and relationships within and between conceptions and practices of mixed methods evaluation through a descriptive process. The selection of cases is therefore purposive and includes fourteen published evaluation articles on projects/programmes. An analytical framework is developed on the basis of a literature review on mixed methods research and background literature on evaluation research. This framework guides the qualitative content analysis of each case study and the cross-case analysis across the fourteen studies to identify common patterns.

The findings reveal two prominent perspectives of mixed methods evaluation prevailing among evaluation practitioners. The first (labeled a 'strong' conception) has the intention of and places emphasis on the integration of the qualitative and quantitative components, with the primary objective of obtaining richer evaluation inferences. In this conception, the use of the methods and the data/inferences thereof are synthesized to achieve this goal. This conception is congruent with mixed methods purposes of: - 'complementarity' and 'triangulation' and is responsive to the 'technical' needs of evaluation. The second perspective (labeled a 'weak' conception) is silent about the integration of the respective methods or data/findings/inferences, qualifying the use of multiple methods and data in a single study as sufficing for a mixed methods approach. It resonates with justifications of mixed methods research that address issues of: - comprehensiveness, multiple view points, inclusiveness and democracy and seems more tailored to the 'political' needs of evaluation. The findings also reveal that the resulting

multiple inferences from this ‘weak’ conception can weaken each other when contradicting or inaccurate qualitative and quantitative findings result, especially when the complimentary function of either method is not planned a priori.

Therefore within the context of realizing richer and more valid evaluation findings/inferences, it is recommended that the purposes and qualification as mixed methods research of the second perspective be re-considered. It is apparent that in embracing the ‘political’ needs of evaluation practice, this conception seems to eschew the ‘technical’ requirements initially intended for a mixed methods approach. This has implications particularly for the mixed methods purpose of ‘expansion’ and rationales of pluralism, inclusiveness and democracy, which are seemingly popular within programme evaluation practice.

# Acknowledgements

I consider this thesis a defining moment to my formal entry into my career passion – monitoring and evaluation of programmes. It is therefore an important and pivotal achievement in my career path and for this, I will always be grateful to the Almighty God who orchestrates everything around my life. I pray that I use this achievement to bless those around me.

I am greatly indebted to my supervisor (Prof. Johann Mouton) for two major things. First, for believing in me and availing an opportunity to pursue a career of my dreams right from the postgraduate diploma through to the Masters programme. Secondly, for the great supervision work he has done on this thesis. This effort has '*aligned*' and sharpened my research and evaluation skills and I will always be grateful.

I am very thankful to the Carnegie Corporation of New York who offered the scholarship for both the postgraduate diploma and the Masters programme. Your support has kept me focused on my studies.

I also extend my appreciation to all the staff of CREST and ERA who have coordinated the postgraduate diploma and masters programmes I've been enrolled for over the last three and a half years. I am particularly grateful to: - Ms. Marthie Van Niekerk, Ms. Lauren Wildschut, and Ms. Charline Mouton. Your coordination role made my studies, numerous travels and stay at Stellenbosch over the years very comfortable and I appreciate your contribution.

To my colleague, classmate, friend and business partner, Komakech Gabriel. First, I thank you for introducing me to the postgraduate diploma. The role you played in this was critical in getting me onboard this career path and I will always be grateful. Secondly, thank you for the company you've been at Stellenbosch and the intellectual engagement that we have had over the years. It has shaped my thinking greatly.

To the management of Makerere University and particularly DICTS, I thank you for the time you have allowed me to pursue my studies. This has been central to ensuring that I complete this thesis in time.

Last but not least, to my honey pie, Anne. Thank you for standing by me all this time, filling in when I've been away, and for believing in my abilities and always being an encouragement. And of course the boys (Jeri and Jona) – the joy you bring has made it much easier.

# Table of Contents

# Chapter 1 - Introduction

## 1.1. Background and rationale

The contemporary evaluation literature advocates for and recommends a pluralistic approach to programme evaluation, with some writers contending that the use of multiple and/or mixed methods for the practice is inevitable. The rationale for such an approach encompasses aspects of both the technical (more valid/truthful, insightful knowledge) and the political (allowing for multiple view points, democratically engaging with difference) requirements of evaluation practice. Its theory of knowledge subscribes to the use of mixed/multiple methods research (MMR) approaches, sometimes termed mixed methods evaluation (MME) as a means of accessing valid and quality data/knowledge, providing better understanding, allowing for multiple, diverse ways of knowing and valuing, hence supporting an all-inclusive/democratic opinion of a social intervention. This perspective to evaluation is underscored by Greene et. al (2001), who with reference to the complexity, dynamism and contextual diversity of the social phenomena studied by evaluators, call for a marshalling of "all multiple ways of knowing…in the service of credible and useful understanding", and recommend development of a mixed-method way of thinking about evaluation.

A number of mixed methods evaluation advocates have reported that mixed methods research has been and is prevalent in contemporary evaluation practices. Greene (1997) refers to the practice as being characteristically pluralistic, embracing diverse perspectives, methods, data and values. Chen (2006) writes that theory-based evaluations "have frequently applied mixed methods in the past". Rallis and Rossman (2003) argue that "…mixed methods designs have been used in evaluation for more than three decades to answer formative, process, descriptive and implementation questions", and also note that "… this pragmatic approach to answering evaluation questions is integral to evaluation practice". Madison (2002) writes, "Evaluators who are concerned more with pragmatics than with competing epistemologies have brought multi- and mixed-method evaluations into common practice". Bleadsoe and Graham (2005) believe that evaluators are more likely to use the elements of multiple evaluation approaches when conducting studies, a practice Greene (2008) accords to the practical demands of the contexts in which they work. Riggin (1997) in endorsing this practice writes, "Evaluators have learned that combining quantitative and qualitative information is not only advisable but inevitable". McConney et.al (2002) add, "…mixed-method rather than mono-method approaches have become firmly established as common practice in programme evaluation".

While the evidence of the use of mixed methods research in evaluation may be very apparent, it is prudent to understand what the notion of mixed methods evaluation is to appreciate this 'overwhelming' use of mixed methods research within evaluation studies. The terminology "mixed methods evaluation" is coined from a combination of the terms "mixed methods research" and "evaluation". In Chapter 2 of this thesis, we present a detailed discussion of the methodology of mixed methods research, providing an in-depth understanding of the approach. Rossi et. al (2004) define Program evaluation as "…the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions". In the following sections, we expound on the concept of evaluation by clarifying and elaborating on three specific issues that elucidate the relationship between social research methods and the process of evaluation. These are: - (i) the nature of 'value claims', (ii) how they are related to the research facts and, (iii) the role of research methods within an evaluation context. There is a need to understand the relationship between the 'results' from the use of social research methods and the process/function of valuing in evaluation. This includes a clarification of what the product of an evaluation study is, what is meant by evaluative inferences or conclusions and in so doing locating how the use of social research methods fits into the pursuit of valuing. To this end, a brief review of the opinions of evaluation 'theorists', particularly those emphasizing the notion of 'valuing' (i.e. discussing the 'how' of valuing) follows in the sections below. Their discussions about how to assess the merit/worth/value of a programme give some insight into this blend of research facts and value claims.

The development of the notion of valuing traces its origins from the early work of Scriven (1967, 1972a) where he emphasizes the central role played by the evaluator in making value judgments, i.e. the evaluator as a valuing agent. He views it as unnecessary to explain why a programme or product works in determining its value. He also advocates for a "goal-free evaluation", in which the goals and objectives of the programme are rejected as the starting points, with preference given to the evaluator having the responsibility of determining which programme outcomes to examine. Premised on his meta-theoretic logic which posits a rule-governed process for drawing conclusions about the merit of the evaluand, he places emphasis on the a-priori determination of performance criteria and standards based on which value is judged. He advocates for an explication of: - criteria, resources, rules, standards, functions,

needs and weights. Scriven (1994) clearly delineates the role of the evaluator, limiting it to the delivery of forthright statements of merit, and cautioning that it is not the evaluator's work to help translate findings into action. With regard to the research-evaluation linkage, Scriven (1991) contrasts evaluation and research noting,

> *"What distinguishes evaluation from other applied research is at most that is leads to evaluative conclusions, and to get to them requires identifying standards and performance data, and the integration of the two".*

Scriven (2003) adds that evaluation is not limited only to the process of determining facts about things and their effects. He argues, "Evaluation must, by definition, lead to a particular type of conclusion – one about merit, worth, or significance". He therefore proposes that evaluations involve three components: - (i) the empirical study, (ii) collecting the set of perceived as well as defensible values that are substantially relevant to the results of the empirical study, and (iii) integrating the two into a report within an evaluative claim as its conclusion. He argues that only the further steps of (ii) and (ii) are what lead to an evaluative conclusion and distinguish an evaluator from an empirical researcher. From this foregoing discussion, it seems that Scriven conceives the process of producing the facts (i.e. the empirical study) and that of valuing as separate entities. However, Scriven (1966, 1983a, 1983b, cited by Shadish et. al, 1991) also notes that empirical facts should inform debates about values, and help decide which values are preferred, and that values can be investigated and justified empirically. He adds that value claims are similar to scientific constructs, with the facts acting as the observable variables. He argues that scientific constructs are not directly observed, but are indirectly observed or inferred from the results of tests. In proposing a logic of evaluation, Scriven (1980, cited by Shadish et. al, 1991) proposes three evaluation activities of: - criteria determination (identifying the dimensions on which the evaluand must do well to be good), setting standards (how well the evaluand must do on each dimension to be good), and measuring performance (measuring the evaluand and comparing the results to the standards).

Scriven's evaluation approach is premised on pre-defining performance criteria/standards, which effectively delineates research facts from the act of valuing. This thinking is shared by Rallis and Rossman (2003) who elaborate more on the specific activities involved in the evaluation process. They argue that evaluation is a specialized form of social science research, noting that "While employing the methods of social science, evaluation has the fundamental purpose of making judgments about the merit and worth of programmes". The pragmatic

framework Rallis and Rossman (ibid) propose entails three primary evaluation activities of: - Description, Comparison, and Prediction. For Description, they note that "…an evaluation describes so as to understand and appreciate the various aspects of the programme or intervention". They add that the role of this component towards judgment of merit/worth has to do with the eliciting of the evaluand's quality through the description. They write,

> *"The details of the picture reveal the programme's inherent goodness or quality – as well as its shortcomings or weaknesses…merit are revealed in the attributes-or intrinsic qualities – of the programme or intervention…Thus, judgments of merit depend on detailed or thick descriptions of these characteristic and attributes, that is, descriptions that allow stakeholders to interpret activities and events".*

Rallis and Rossman (ibid) therefore consider the need to avail the most revealing picture, the fullest and thickest description about the evaluand as one of the rationales for drawing on a mix or multiplicity of methods. They additionally write that evaluation rarely ends with description but moves beyond to, "…use comparisons for judgment in relation to another service or intervention or to a standard" and that this is more salient at the data analysis stage than in data collection. They explain,

> *"The comparison activity is especially important in evaluation because judgment about a programme's merit in relation to a standard…are often more powerful to stakeholders than judgments based on intrinsic quality".*

They relate the Prediction dimension to the worth of a programme and note that prediction tries to "…make judgments about the overall value of the programme to some group of constituents or to society in general, including recommendation about future programming and funding".

This emphasis on the definition of criteria/standards is also shared by House (2004), whose contribution is more towards the process of determining the criteria/standards. He advocates for what he terms 'deliberative democratic evaluation' that uses concepts of democracy to arrive at justifiable evaluative conclusions. It has three principles of inclusion, dialogue and deliberation. These three principles ensure that: - all relevant interests of stakeholders are captured; that the 'real' as opposed to the 'perceived' interests of stakeholders are teased out; and that facts and value claims are examined through rational processes rather than taken as given. He argues that evaluators should concentrate on drawing conclusions, a process that entails employing appropriate methods as one component, but also discovering the right criteria and standards of

comparison and addressing the criteria and standards in the data collection, analysis and inference to conclusions.

An alternative valuing dimension has emerged which is a move away from Scriven's rubric *(the rule governed approach of a-priori criteria and standards, description, then judgment)* to a more dialectic hermeneutic approach of describing and judging an evaluand. Its central tenet is that valuing is experiential, making it inseparable from the activity of describing. This notion is elaborated upon by Shadish et. al (1991) who note that the starting point of constructing criteria of merit is an understanding of the evaluand. They interpret Scriven's statement,

> *"Once one understands the nature of the evaluand…one will often understand fully what it takes to be a better and a worse instance of what type of evaluand",*

as implying that criteria of merit (a valuing product) stem from descriptors of the evaluand and are therefore subsets of evaluand description. Below we elaborate on this 'facts' and 'values' relationship from a discussion of other values-oriented evaluation theorists

Eisner (1994, cited by Alkin, 2004) rejecting the extensive use of research models employing experimental and quasi-experimental designs argues that things that matter cannot be measured quantitatively. He contends, "Evaluation requires a sophisticated, interpretive map not only to separate what is trivial from what is significant, but also to understand the meaning of what is known". His theoretic views are premised on two notions of connoisseurship and criticism, which he proposes as attributes/competencies of the evaluator. He describes connoisseurship as "the art of appreciation…involving, the ability to see, not merely to look". He adds, "To do this we have to develop the ability to name and appreciate the different dimensions of situations and experiences, and the way they relate one to another. We have to be able to draw upon, and make use of, a wide array of information". He defines criticism as "the art of disclosure", approached as the process of enabling others to see the qualities of something (Eisner, ibid). Eisner thus considers as central the valuing role of the evaluator, one that goes beyond the competence required with Scriven's logic of a-priori criteria and standards, but that which is experiential and phenomenological, relying exclusively on qualitative methods.

Stake (2004) argues that "…seeing and judging the evaluand regularly are part of the same act and that the task of evaluation is as much a matter of refining early perceptions of quality as of building a body of evidence to determine level of quality". This thinking diffuses the logical

demarcations defined in Scriven's rubric. Stake, et. al (1997) build on this argument, noting that what evaluators do is, "…more a matter of seeking to understand what is going on and devising representations of production, performance, effectiveness, crisis management, staffing, etc that help describe the evaluand". While agreeing that judgment is the essential logic of evaluation, they have little faith in rubrics for doing so. They however do not disregard the role of rubrics, noting that explication and rules help focus the important considerations of an evaluation, keeping the evaluator from overlooking important ingredients. Their point of contention is the limitation that the rubric's criterial treatment of an evaluand has in transforming experiential knowledge of it into knowledge of selected characteristics, arguing "…there are no representations that mirror reality, none that draw us closer than experience to the real world". They conclude that "It is the human, value-edged, perceptual response to stimulation, to the evaluand's being or doing, that is the essence". They therefore propose a hybrid of both essentialist and relativist thinking, referring to evaluation as 'eclectic thinking', a "…shifting back and forth between the formal and informal, the general and the particular, the hunch and the habit…". In summary, Stake. et.al (ibid) propose to enhance the process of explicating an evaluand's value through experiential knowledge, which they consider complementary to the criterion-based approach. In doing so, they consider the demarcations of a standard evaluation logic blurred, with valuing happening iteratively and dialectically as part of the process of describing the evaluand's quality.

Stake and Schwandt (2006) develop the foregoing argument further in their explication of the notion of quality (merit, value, worth, significance). They are concerned about the criterial approach, arguing that it is the "explicit and sole equation of quality with *performativity* – the measurement of performance against indicators of target achievement", cautioning that this can lead to a substitution of quality with performance. They introduce two concepts of "Quality-as-Measured" and "Quality-as-Experienced" in working towards a solution. In the former, the appraisal of quality is based on clearly articulated and explicit criteria and standards, "where quality is regarded as measurable, and judging quality…has the explicit comparison of the object in question to a set of standards for it". They note the limitations of this conception as the inability of a few indicators at fully representing an evaluand, which is usually more complex. For the latter concept, quality is conceived as a "…phenomenon that we personally experience and only later make technical…" However, they note the limitation of 'quality-as-experienced' as being contingent on the "acuity and credibility of the observer". They also cite the issue of the ability to experience a programme, what they refer to as "embracing quality". They contend that

an evaluand on a small scale is ‘embraceable’ and the evaluator can become experientially acquainted with it and therefore perceive experiential quality. However, when the evaluand is extensive, the evaluator cannot easily embrace it and typically abstracts an evaluand’s quality with criteria and standards. This contrast explicates the reality for the need of both ‘measurement-based’ and ‘experiential-based’ quality measures. They surmise that quality constructs provide essential intellectual structure for disciplined inquiry into quality. However, to effectively explicate quality, there is need for experiential thinking aimed at amplifying and redefining them during the course of the study.

This discussion elucidates two principal concepts which values-oriented theorists conceive as approaches to “evaluation research”: - a criterion/standards based approach and an experiential dialectic approach. Within the criterion based approach, describing and valuing a programme are separate activities, are linearly and logically linked, happening in succession, and involving clear criteria and standards for comparison in making an evaluative judgment. Variations may exist in the way the criteria/standards are established. One central tenet of this approach is a comparison of the evaluand description to the pre-defined standards along the criteria of assessment. Therefore, the description phase while possibly including elements of programme value or merit does not suffice as an evaluative conclusion. The programme merit/worth is only elicited though comparison with the pre-defined standards for the respective criteria. On the other hand, the experiential approach conceives describing and valuing as a dialectic process, with valuing seen as part of the description process. Facts are representations of values which guide inquiry, and the quality or richness of this factual representation of quality is sought to elucidate the merit/value/worth of an evaluand. While the former approach is explicit about comparison to pre-defined criteria/standards, this latter approach infers a means of comparison to some implicitly defined standards determined through the experiential knowledge of the valuing agent. It is however noteworthy that a common theme shared by both evaluation research approaches is the need to compare an evaluand’s description to some criteria/standard, with the variation being in the way this standard is established.

From the discussion thus far, the role of research methods in evaluation for either approach is more-or-less similar, with a slight but overlapping variation in the purpose for which the description is done. A further concern for evaluation theorists with regard to this evaluand description has focused on two issues: - (i) how it can be attributed to the programme, and (ii) how it can be generalized beyond the study sample. We briefly discuss opinions of evaluation

researchers on these two issues in the following sections to get a deeper insight into the nature of inferences in programme evaluation.

## 1.2. Causal inferences in programme evaluation

Evaluation studies typically focus on three programme aspects: - the design (efficacy), the implementation (efficiency), and the outcomes/impact (effectiveness). While the merit of the first two programme aspects is directly elicited through the evaluand description, determining the merit of the last aspect (effectiveness) varies. The description of effectiveness typically focuses on the change in that (e.g. target group) which the programme was expected to influence. Hence, to impute the described effects to the merit of the programme, there is a need to attribute or causally relate these effects to the programme '*effort'*. This issue has been termed 'causation' or 'causal attribution' and is a central guiding principle in the design of evaluation research studies. In the next few paragraphs, we present a discussion on how various researchers have approached this issue.

Causality is synonymous with 'Internal validity' whose key question Trochim (2006) notes as "…whether observed changes can be attributed to your programme or intervention (i.e., the cause) and not to other possible causes *(sometimes described as "alternative explanations" for the outcome)*". For evaluation, the early approach to causality was the use of programme goals to formulate causal hypotheses, which would then be tested using experimental approaches. Theorists like Chen (2004) promote the experimental approach by proposing to supplement it with evaluation theory. In his development of the concept of theory-driven evaluation, he is concerned about the experiment's failure to provide any explanations for the success or failure of a programme. He proposes that experimental approaches "… should be used in conjunction with a priori knowledge and theory to build models of the treatment process and implementation system to produce evaluations that are more efficient and that yield more information about how to achieve desired effects" (Chen & Rossi, 1983).

Lately, there has been increasing emphasis on non-causal issues and questions about causal explanation in evaluation. The need for explanatory knowledge has become more prominent, with an emphasis on explaining effects as opposed to just describing them. The advent of qualitative research advocating for explanatory knowledge has been promoted as a prominent alternative. Researchers like Maxwell (2004) coming from the qualitative paradigm have contributed to this development with a realist approach to causal explanation. He argues that

"realists typically understand causality as consisting not of regularities but of real…causal mechanisms and processes, which may or may not produce regularities". Using Mohr's (1982, 1996) labels of 'variance theory' and 'process theory', he contrasts variance theory *(which deals with variables and the correlations among them, and is mainly associated with quantitative methods)* with 'process theory' *(which deals with events and the processes that connect them)*. He argues that 'process theory' is less amenable to statistical approaches and is more tailored towards in-depth studies of a few cases. It is on this premise that he justifies qualitative causal explanation. However, critics of the qualitative causal explanation approach note its limitations regarding addressing the counterfactual issue. They particularly argue against the 'thick description' of case studies and theory-based evaluation as proposed alternatives to experiments. Cook and Shadish (1986) write, "…qualitative methods usually produce unclear knowledge about the counterfactual…how those who received treatment would have changed without treatment". They however observe that the combination of case studies with experimental design can improve the causal inference through the inclusion of designs like comparison groups and pre-treatment observations. They advocate for a combination of qualitative methods within experiments to give more value when substantial uncertainty reduction about causation is required. Regarding theory-based evaluation, they note its limitations for strong causal inferences when testing causal hypotheses. These are mainly premised on two issues: - the non-clarity of most theories, which could be interpreted in diverse ways and the linearity of theory flow, which omits reciprocal feedback or external contingencies that might moderate the entire flow.

Davidson (2004) raises the issue of the level of certainty required by the client regarding the need to demonstrate causation. He notes that it is important to be clear upfront about the level of certainty required because each decision-making context requires a different level of certainty. He also notes the importance of identifying rival explanations *(also dependent on the certainty level required)* for making stronger and more defensible conclusions. He recommends a blend of strategies for addressing the causation issues. He argues that these are mostly 'commonsense' approaches like: - asking observers (e.g. beneficiaries); checking whether the content of the evaluand matches the outcome; looking for other telltale patterns that suggest one cause or another; checking whether the timing of outcomes makes sense; checking whether the "dose" is related logically to the "response"; and identifying and checking the underlying causal mechanism(s). The more scientific ones are quantitative and include: - making comparisons with a "control or "comparison" group and controlling statistically for

extraneous variables. Alluding to a preference for the less scientific approaches for evaluation, Cronbach (1982) asserts, "…potential users of evaluation are less concerned than academics with reducing the final few grains of uncertainty about knowledge claims; that prospective users are more willing to trust their own experience and tacit knowledge for ruling out validity threats; and that they also expect to act upon whatever knowledge base is available, however serious its deficiencies". The preference is for generating many findings, even at the cost of achieving less certainty about any one of them (Cook and Shadish, 1986).

### 1.3. Generalizing inferences in programme evaluation

It is typical in evaluation research to require the evaluand description of merit/worth to be representative of a population bigger than or different from the study sample. This has been an issue of concern within the evaluation domain and we discuss opinions of various evaluation theorists in this section. In the context of evaluation, generalizability is defined using the question, "*Can the programme be used with similar results if we use it, with other content, at other sites, with other staff, with other recipients, in other climates (social, political, physical), and so on"* (Scriven, 2005). It is also synonymous with the terms 'external validity', which is a question of, *"to what population settings, treatment variables, and measurement variables a cause-effect relationship can be generalized"* (Campbell, 1966; Campbell, 1957 cited by Cook, 2005). Central to this issue is the concept of the 'population' to which the sample is being generalized. A number of authors have discussed this for the evaluation context. Campbell (ibid) considers two populations: - an almost unique population from which the sample is extracted and then the infinitely large universe. He argues that evaluation can only realistically generalize to the former and not the latter. This early conception of generalization relates to the scientific approaches of sampling. Generalizability as conceived within the experimental approach in which sampling with known probability from some clearly designated universe was the initially preferred technique. However, critics of the experimental approach note the difficulties in defining, "some types of universe, particularly when historical times or physical situations are at issue" (Cook and Shadish, 1986). They also cite "…the variability between projects, and between clients and practitioners within projects" arguing that this "requires that samples have to be "large" and hence more expensive if formal representativeness is to be achieved within "reasonable" limits". They add that "…formal probability sampling requires specifying a target population from which sampling then takes place, but defining such populations is difficult for some targets of generalization such as treatments".

Subsequently, other authors introduced the notion of extrapolating the sample results to the population of interest. Cronbach (1982) disregards the 'Universe' population, and focuses on 'extrapolating' the sample results to what Sasaki (2005) terms a 'policy-target' population. He does this by identifying particular instances in the sample that closely manifest this population. Stake (1995) also eschews the 'Universe' population but proposes 'naturalistic evaluation', for which he contends that every case study has some unique information to possibly modify *(effectively generalizing)* the already made generalizations of evaluators. Cronbach (1982) introduces an alternative perspective to generalization, emphasizing that it is also a product of causal explanation. He argues, "The more we know about the plethora of contingencies on which programme or project effectiveness depends, the more likely it is that we will be able to transfer successful practices to other sites that have not yet been studied". He proposes that generalization could be "attained by extrapolating through causal explanation, either using causal modeling or the "thick description" of qualitative methods" (Alkin and Christie, 2004). These qualitative approaches to sampling are becoming, particularly purposive sampling approaches "…that emphasize selecting instances from within a population that is either presumptively modal or manifestly heterogeneous" (Cook and Shadish, 1986). The rationale for the modal cases is to ascertain whether causal relationships can be generalized to the most frequently occurring types of persons or settings. The rationale behind the heterogeneous cases is to test whether the causal relationships/hypotheses posited will remain valid under differing persons and settings.

These discussions around generalization illustrate an increasing appreciation of the more qualitative *(causal explanation and transferability)* approaches to addressing the concerns of causation and generalizability in evaluation. It is apparent that qualitative explanations that logically link a programme's activities to the stated effects and provide lessons for applicability to other contexts are becoming equally prominent and acceptable in addressing these two concerns within evaluation practice.

The review of the evaluation literature thus far gives an insight into the research product *(i.e. the nature of inferences or conclusions)* sought in evaluation studies. The three particular features of these inferences that have been identified are: - (i) the notion of valuing as including the evaluand description and a need for comparison to some criteria and standards in determining programme merit/worth, (ii) the need to causally link the programme 'effort' to the effects, and (iii) the need to represent the inferences to other populations/contexts. In the following three

sections, a review of how mixed methods research has been discussed within the context of evaluation, and particularly in the making of evaluative inferences is presented to establish the progress so far made with this 'novel' evaluation approach.

A number of researchers have explored the notion of mixed methods research from an evaluation context. One of the early and pre-dominant mixed methods research conceptual frameworks (Greene, et. al, 1989) is based on an analysis of fifty seven empirical mixed methods evaluations. This framework elicits mixed methods research purposes and designs that evaluators use in practice and has to a large extent informed mixed methods research designs in the social and other sciences. Others like Caracelli and Greene (1993) have proposed four integrative data analysis strategies for mixed methods evaluation designs derived from and illustrated by empirical practice. They discuss the appropriateness of these strategies for different kinds of mixed methods intents. Caracelli and Greene (1997) have also proposed ways of creating mixed methods evaluation designs and presented two broad classes: - component and integrated designs. Miller and Fredricks (2006) have explored the relevance of mixed methods research to educational evaluation and argued for a particular form of mixed-methods design (quantitative-dominant sequential analysis) as proving useful for some educational evaluation and policy studies.

Within the context of theory driven evaluations, Chen(2006) writes that the comprehensive scope of theory driven evaluations involves the sequential combination of its two primary tasks of : - (i) facilitating stakeholders in clarifying or developing their programme theory, and (ii) empirically assessing programme theory and proposes four strategies for using mixed methods. These include: - the Switch strategy in which one first applies qualitative methods to clarify stakeholders' programme theory and then uses quantitative methods to assess the programme theory; the Complementary strategy which involves the use of qualitative and quantitative methods to collect different pieces of information for assessing a programme theory in order to gain clear understanding of a programme; the Contextual overlaying strategy which refers the use of a method (quantitative or qualitative) to collect contextual information for assisting in interpreting the data or reconciling inconsistent findings; and the Triangulation assessment strategy where multiple or mixed methods are applied in cross-validating an observed phenomenon.

A few authors have explored the issue of the benefit of integrating or mixing methods in evaluation. Madey (1982) writes that the integration of quantitative and qualitative methods

within a single evaluation has synergistic effects in the three major research phases of design, data collection and analysis. Her emphasis is on how the different methods enhance each other. Illustrating with a specific evaluation study, she demonstrates how qualitative methods can enrich quantitative designs by improving both the sampling framework and the focus of the overall evaluation design. Similarly, she illustrates how quantitative techniques can contribute to qualitative methods by: - identifying both representative and unrepresentative cases during sampling; using the quantitative results to provide leads to further interviewing; focusing the study on overlooked respondents and correction of the elite bias during data collection; and correction of the 'holistic fallacy' and verification of qualitative interpretation during data analysis. Greene, et. al (2001) illustrate the concept of 'better understanding' of social phenomena resulting from the use of mixed methods with case examples through which they demonstrate the following perspectives: - 'Enhanced validity and credibility of findings' through triangulation in which different methods ideally with offsetting biases are used to measure the same phenomenon, effectively ruling out the threat to validity; 'Greater comprehensiveness of findings' where the lenses of different methods are focused on different aspects of a phenomenon to provide a more complete and comprehensive account of a phenomenon; 'More insightful understanding' where non-convergent or conflicting results lead to new insights and hence further explorations about the phenomenon; and 'Increased value consciousness and diversity'.

The foregoing discussion illustrates that research on mixed methods evaluation has proposed a number of prescriptions about how the qualitative and quantitative methods can be integrated, directed by a research design that is linked to a research purpose. Some of these studies have been based on descriptions of evaluation practices. A few have gone further and addressed the issue of the potential benefits of mixing the methods, mainly emphasizing how the different methods enrich each other for better research design towards more valid results. However, none of the literature reviewed has discussed mixed methods evaluation within the context of the product of the primary intent for using mixed methods in evaluation studies, i.e. the making of richer conclusions or inferences about programme merit/worth. Additionally, there is a dearth of literature that describes and relates the understanding and use of mixed methods evaluation to the conclusions/inferences made from a practice perspective. These are the issues that constitute the research problems of this study.

## 1.4. Research problem

As illustrated from the brief review of the foregoing literature, the role of mixed methods research approaches towards richer evaluation findings cannot be over-emphasized. Additionally, the pragmatic, democratic and political facets of the evaluation practice call for a pluralistic methodological approach, which mixed methods research proffers. It has been clarified that methods need to be combined or integrated for a particular purpose or towards a particular end. The brief review of the literature on valuing underscores and focuses this need in its illustration of the peculiar nature of inferences that may be considered important, valid and relevant within an evaluation context. This forms the background against which the research problem is developed. The main research question that has guided the thesis is: -

***How has the notion of mixed methods evaluation been understood and implemented by evaluation practitioners?***

## 1.5. Objectives of the Research

The overarching aim of this thesis is to establish how the perception and implementation of mixed methods research among evaluation practitioners influences the nature of conclusions/inferences they make. The specific objectives include: -

1. To get an in-depth understanding of the methodology of mixed methods research through a detailed review and analysis of the related literature.
2. To elicit evaluation practitioners' understandings and uses of the approach by establishing: -
    a. The justification(s) given for a mixed methods research approach and how these guide the actual research implementation.
    b. The different ways the qualitative and quantitative methods are defined, used and integrated.
    c. The nature of evaluation findings/inferences made and how (if at all) the use of the qualitative and quantitative methods is harnessed in this respect.

## 1.6. Research design and Methodology

The thesis focuses on reviewing a particular social research methodology and exploring how it is approached within evaluation study contexts. A methodological study design is therefore most suited for this purpose. Mouton (2001) defines a methodological study as that "…aimed at developing new methods of data collection and sometimes also validating a newly developed

instrument through a pilot study". While this definition gives emphasis to methods, it has been appropriated to address the need of the methodological research question posed in this thesis.

This thesis aims at identifying patterns and relationships within and between conceptions and practices of mixed methods evaluation through a descriptive process *(i.e. identifying and describing emerging issues/themes/trends of specific MMR approaches and designs).* The selection of cases is therefore purposive and the choice of cases to study is guided by an aim of including an appropriate and adequate number of studies representing the various mixed methods research attributes like: - purposes/rationales for mixing, the mixed methods research designs, data types among others.

The cases for analysis are published evaluation articles on projects/programmes. A selection of fourteen evaluation studies is used as the source of data for the study.

An analytical framework emerging from the review of the literature on mixed methods research is used to guide the study. A qualitative content analysis of each evaluation study and a cross-case analysis across the studies are carried out to identify common patterns/themes. Specifically, the following aspects of the study are considered: - the programme context, the rationale for mixing the methods, the form and use of the qualitative and quantitative components of the study, and the nature of inferences made.

## 1.7. Lay out of the thesis

This first chapter gives the background and rationale, building a case for studying this research topic. In the next chapter, a review of mixed methods research is presented, including aspects of: - the historical developments, its philosophy, definitions of mixed methods research, the various designs, and criticisms. This review provides an understanding of the issues that have been discussed in mixed methods research and largely informs the analytical framework adopted for the study. The analytical framework emerging from the literature review is used as a basis for the content and thematic analysis of the selected cases. The results and discussion thereof are presented in Chapter 4 with the conclusions and recommendations coming last in Chapter 5.

# Chapter 2 – A review of the mixed methods research approach

Mixed methods research (MMR) is the contemporary social inquiry approach that has taken center-stage in recent research methodology discussions. It has been proposed and preferred as a solution to the paradigm wars in addition to providing more valid, quality and richer results and inferences as compared to the traditional mono-method approaches. While not yet fully developed like its methodological paradigm peers of qualitative, quantitative and Participatory Action Research, much research has been carried out to position it at a level where is could be considered a fully recognized and distinct methodological paradigm. In this chapter, the various developments of the approach are explored to elicit a clear understanding of how researchers have conceptualized MMR. Aspects of this approach that are reviewed include: - a trace of its historical development; the philosophical assumptions that undergird the approach; how researchers have defined MMR; and the different aspects of the MMR methodology including MMR questions, sampling, design, analysis, and validity. Through this review, an analytical framework emerges to guide the empirical study.

## 2.1. History of the development of mixed methods research

This discussion is informed by the work of two authors who discuss the history of MMR. Creswell, et. al (2007) review a sketch of the history of MMR by Tashakkori and Teddlie (1998) and organize it into four time periods. These include the 'Formative'; The 'Paradigm debate'; the 'procedural development'; and the 'Advocacy as a separate design'. For each of these periods, Tashakkori and Teddlie (ibid) identify important writers and their contribution to the development of MMR. They describe the 'Formative period' (1950s – 1980s) as characterized by the initial interest to use more than one method in a single study, making specific reference to writers who advocated for the collection of multiple forms of quantitative and qualitative data and those who combined both qualitative and quantitative data in their studies. The 'paradigm debate' period (1970s – 1980s) is defined as starting with the qualitative researchers' insistence that since different assumptions provided the foundations for qualitative and qualitative research, their combining was untenable. Creswell et. al (1998) note that subsequent writers challenged this position, with the eventual classifications of researchers as purists, situationalists and pragmatics depending on their opinion about the combination

of paradigms, use of methods and paradigms in addressing research problems. They add that although the issue of reconciling the paradigms is still apparent, calls have been made for an alternative paradigm (pragmatism) and ways of engaging the two paradigms for mixed methods research. Creswell et. al (ibid) describe the third period ('Procedural developments' - starting in the late 1980s into the 1990s) as a shift towards the methods or procedures of designing a mixed methods study despite the ongoing paradigm debate. They note the premising on Greene. et. al's (1989) empirical study which proposes six classifications of MMR designs. They also refer to other researchers who, following in the footsteps of Greene et. al (1989), contribute to this discussion in various ways. The issues researchers discuss during this period include: - linking multi-method research in the various steps of a research process; ways of implementing the different quantitative and qualitative components of a study; developing of specific types of mixed methods designs; choosing among various designs; and issues of validity and making inferences. For the last period ('Advocacy as separate design' – 2003 to-date), they cite developments that show indications of interest towards establishing it was a unique research methodology. A prominent 'landmark' they refer to is the handbook of mixed methods (Teddlie & Tashakkori, 2003a) with many chapters solely devoted to discussions on various issues and in many disciplines. They also cite authors (Johnson and Onwuegbuzie, 2004) who advocate for the consideration of MMR as a distinct methodology alongside quantitative and qualitative approaches. Other indicators included are: - the inclusion of mixed methods in research guidelines; workshops that include discussions on mixed methods; workshops on mixed methods; journal articles on mixed methods studies; special interest groups; and its increasing application in different research disciplines. They conclude by noting a cross-cultural, interdisciplinary, publication, private and public funding interest for MMR as additional proof of the ascendency of the methodology.

Teddlie and Tashakkori (2003) present a historical analysis of the emergence of mixed methods. They map it onto Denzin and Lincoln's (1994) five 'moments' in qualitative research and describe the developments in MMR along these periods. They write that during the 'traditional' period (1900-1950), there was substantial mixed methods research but without any methodological controversies. This was despite some debates about the relative merits of either qualitative or quantitative research. They specifically cite two major studies and note that interviews, observations and experimental studies

were used in one of them in particular. They add that though the distinct field of mixed methods research had not emerged, research designs emerged that began to be called "multi-method" or "mixed" during the next period (1950-1970s). They cite three studies in the field of psychology in which the use of mixed methodologies occurred. They refer in particular to Campbell and Fiske's (1959) "multi-trait multi-method matrix" as the first explicit multi-method design, inevitably leading to studies that mixed quantitative and qualitative methods. They combine the two qualitative periods of "blurred genres" (1970 – 1986) and "Crisis of representation" (1986 – 1990) into one which they term "the ascendance of constructivism, followed by the paradigm wars". They discuss two major developments during this time: - the earlier period focused on triangulation, with specific reference to Denzin (1978) who introduced the term and discussed different types of triangulation. They also cite Jick (1979) who discussed how the weaknesses of one method are offset by the strengths of another with specific reference to "across methods" triangulation. They refer to the last period (1990 – present) as the "emergence of pragmatism and the compatibility thesis". They refer to Howe's (1988) advocacy for pragmatism as the philosophical paradigm for mixed methods and a number of seminal works aimed at establishing mixed methods as a separate field. These seminal works focused mainly at: - typologies of mixed methods designs, key words and definitions and different paradigm formulations.

From these two descriptions, four overlapping stages are common in the development of mixed methods research. The first are what authors term the classic 'mixed methods' studies which were evident from the early to the middle part of the century (1939 - 1961); the next is a proliferation of multiple methods designs/triangulation (1959 – to-date); the third is the mixed methods 'movement' (1985 – to-date) starting with the philosophical paradigms and progressing into establishment as a distinct research methodology; and the last are mixed methods designs (1989 – to-date). These stages are explored in detail in the subsequent sections to establish how they have evolved and influenced the development of the field.

### 2.2. The classic 'mixed methods' studies

The earliest classic studies using 'mixed methods' that have been cited are the Hawthorne effect experiments (1939) and the Yankee city studies (1941). Since then, mixed methods have been employed in a number of studies without necessarily being

formally labeled as mixed methods. A review of three of these studies is presented to provide a backdrop against which the peculiarities (if any) of the formal mixed methods 'movement' can be compared.

***The Hawthorne studies***

The Hawthorne studies (Roethlisberger. F. J and Dickson. W. J., 1939) comprised a long series of investigations into the importance for work behavior and attitudes of a variety of physical, economic, and social variables. The principal investigations were carried out between 1927 and 1932. They had five stages viz: Stage I: The Relay Assembly Test Room Study (New incentive system and new supervision); Stage II: The Second Relay Assembly Group Study (New incentive system only); Stage III: The Mica Splitting Test Room Study (New supervision only); Stage IV: The Interviewing Program; and Stage V: The Bank-Wiring Observation Room Study. Stages I to III constituted a series of partially controlled experimental studies which were initially intended to explore the effects on work behavior of variations in physical conditions of work, especially variations in rest pauses and in hours of work, but also in payment system, temperature, humidity, etc. Stages II and III were designed to check on the Stage I conclusion. Stage IV was an interviewing program undertaken to explore worker attitudes. Stage V was a study of informal group organization in the work situation. The two later studies (IV and V) resulted directly from conclusions based on Stages I-III about the superior influence of social needs. Observations made in both were interpreted in the light of such prior conclusions. These studies demonstrate the early use and integration of multiple quantitative and qualitative methodologies to study the same phenomenon, with the qualitative methods complementing the primary quantitative studies.

***The end-of-world cult***

The end-of-world cult (Festinger, et. .al, 1956) was a psychology study of an 'End-of-the-world cult' and the consequences of cult members of the failure of its predictions. The study began with a variable-oriented theory and a hypothesis about the conditions under which disconfirmation of belief will paradoxically be followed by increased commitment. The data were collected entirely through participant observation by a number of researchers pretending to be cult converts. This called for intensive involvement of the researchers in the cult activities. The cult members were categorized into two groups based on two independent variables: - degree of prior commitment and social support.

The experiment involved the comparison of results from the two groups. The observational data were analyzed quantitatively to compare the results of the two groups. This study can be summarized as a quasi-experiment, with pre- and post-intervention qualitative data collection and a comparison of two parts of the group.

***The Robber's cave experiment***

The Robbers Cave experiment (Sherif. et. .al, 1961) on inter-group conflict and co-operation was established as an interdisciplinary "psychological" and "sociological" approach to the testing of a number of hypotheses about inter-group relations. Twenty two boys were selected and were divided by the researchers into two groups with efforts being made to balance the physical, mental and social talents of the groups. During the first five or six days each group was given a series of activities which encouraged the members to develop a common bond. The two groups solidified their identities and in each case spontaneously took on a name. After the first few days, the researchers, playing the roles of camp staff initiated the second stage of the experiment. They arranged a series of competitive activities where the winning group members received attractive awards and the losing group members did not receive anything. They observed increasing hostility between the two groups. The next stage of the experiment consisted of a number of meet-and-greet activities which were designed to provide reconciliatory opportunities. In the final stage of the experiment the two groups were placed in situations where there was a compelling super-ordinate goal which could not be achieved by one group acting alone. What stands out for this experiment is the extensive use of the qualitative participant observation methods as a means of data collection for the experiment.

These three studies illustrate the use of some variants of mixed/multiple methods long before the advent of the mixed-method 'movement'. The first is a use of multiple QUAL and QUAN methodologies in a complementary way at different stages of the study, with the QUAL methodologies further exploring discoveries from the QUAN components. This study illustrates the co-existence of multiple paradigms within a single study. The latter two are largely experiments using qualitative data collection methods with quantitative analysis, illustrating the use of QUAL and QUAN methods guided within a single paradigm inquiry framework.

## 2.3. Multiple methods designs and triangulation

The early 'formal' discussions following the classic experiments focused on the use of multiple methods designs in research. The prominent authors of this period are discussed in the following sections.

Campbell and Fiske (1959) proposed the first multiple method design termed a multi-trait multi-method (MTMM) matrix which used more than one quantitative method to measure a psychological trait. Teddlie and Tashakkori (2003) note that the purpose was to ensure that the variance in the research findings was accounted for by the trait under study and not by the method that was employed to measure it. Ferketich, et. al (1991) add that the basic underlying tenets of the MTMM matrix are that: - tests designed to measure the same construct should correlate highly among themselves and that tests measuring one construct should not correlate with tests measuring other constructs. Thus, based on the first tenet, convergent validity is supported by the presence of relatively strong correlations among measures of the same construct; and based on the second tenet, discriminant validity is supported by the presence of relatively small correlations among tests measuring other constructs regardless of the method used. Johnson. et. al (2007) clarify that this idea of 'multiple operationalism' is more of a measurement and construct validation technique, in its original formulation, than it is a full research methodology. They add, "…early researchers' idea of multiple operationalism follows more closely what today is called *multimethod research*, in contrast to what currently is called *mixed methods Research"*. They however note that Campbell and Fiske (1959) are rightfully credited as being the first to show explicitly how to use multiple research methods for validation purposes.

Webb, Campbell, Schwartz and Sechrest (1966) extended the ideas of Campbell and Fiske (1959) by placing more emphasis on what is being measured as opposed to validating the methods used. They suggest that "Once a proposition has been confirmed by two or more independent measurement processes, the uncertainty of its interpretation is greatly reduced. The most persuasive evidence comes through a triangulation of measurement processes. If a proposition can survive the onslaught of a series of imperfect measures, with all their irrelevant error, confidence should be placed in it". Johnson. et. al (2007) note that Webb et al. (1966) are credited with being the first to coin the term *triangulation*, of the type referred to as between- or across-method

triangulation. The notion of triangulation was broadened by subsequent authors to include dimensions beyond the methods used and the constructs measured. Below we present views of a few of these researchers.

Denzin (1978) develops the concept of triangulation further by classifying it into four basic types according to the focus of: - data, investigator, theory and methodology. He further divides methodological triangulation into “within method” and “between or across methods” triangulation, depending on whether the methods belong to the same or different methodological approaches. He defines data triangulation as using several data sources, for example the inclusion of more than one individual as a data source. Denzin (ibid) broadens the notion of data triangulation to include time and space based on the assumption that understanding a phenomenon requires its examination under a variety of conditions. He defines Investigator triangulation as involving the use of multiple researchers in an empirical study. Acknowledging that a research process typically involves more than one researcher, the issue he considers problematic is who these researchers should be and what their roles should be in the research process. He defines methodological triangulation as involving the use of multiple methods in the examination of the same phenomenon. He suggests that the within- method triangulation approach has limited value, because “….essentially, only one method is being used, and finds the between-methods triangulation strategy more satisfying” (Mathison, 1988). He argues that “…the rationale for the between-methods strategy is that the flaws of one method are often the strengths of another; and by combining methods, observers can achieve the best of each while overcoming their unique deficiencies”. Denzin (1978) also defines Theory triangulation as involving using more than one theoretical framework in the interpretation of the data. However, Denzin does not emphasize this type of triangulation, noting that “…sociologists committed to a given perspective will probably not employ theoretical triangulation”. His inclusion of it is that it underscored the fact that every study is conducted with some theoretical perspective and that this was important for the theoretically uncommitted researchers and for those areas characterized by high theoretical incoherence.

Jick (1979) arguing that the strong advocates for triangulation fail to indicate how it was actually performed seeks to demonstrate how it is accomplished in practice. He first

elaborates on the concept of triangulation, viewing it on a continuum that ranges from simple to complex designs (fig. 2.1).

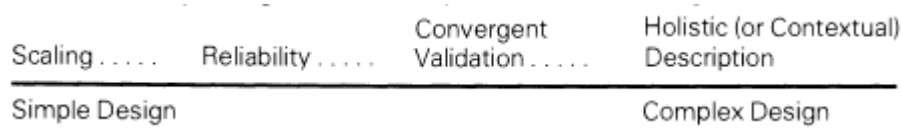


Fig. 2.1. A continuum of triangulation design

At one end (the simple) is what he terms Scaling, i.e. the quantification of qualitative measures; next to it is a more sophisticated triangulation design : - the "within-methods" strategy for testing reliability; next in the continuum is the conventional form, the "between methods" approach designed for convergent validation. He argues that triangulation, could be something other than scaling, reliability, and convergent validation. That it could also capture a more complete, holistic, and contextual portrayal of the unit(s) under study. He contends that it is here that qualitative methods, in particular, could play an especially prominent role by eliciting data and suggesting conclusions to which other methods would be blind. He illustrates the triangulation strategy in a study he conducted on the effects of a merger on employees using various techniques (self- reports, interviews, co-worker observations, data collected through archival sources and unobtrusive measures). He argues that these various techniques and instruments generate a rich and comprehensive picture of the phenomenon of inquiry. Jick (ibid) also establishes that while the various methods together produce largely consistent and convergent results, there are also some surprises and discrepancies in the multi-method results which lead to unexpected findings. He uses such discrepancies to initiate further inquiries to explain or reconcile the disagreement, yielding richer findings. He concludes that "…the process of compiling research material based on multi- methods is useful whether there is convergence or not. Where there is convergence, confidence in the results grows considerably. Findings are no longer attributable to a method artifact. However, where divergent results emerge, alternative, and likely more complex, explanations are generated".

Other researchers proposed structured approaches for triangulation designs. Rossman and Wilson (1985) expand the thinking around the purpose of combining methods from triangulation/confirmation/corroboration to other alternatives. The first is using methods combinations to enable or develop analysis to provide richer data. The second is the use

of methods combinations to initiate new modes of thinking by attending to paradoxes that emerge from the two data sources.

Morse (1991) proposes two types of methodological triangulation - simultaneous or sequential. She defines simultaneous triangulation as the simultaneous use of qualitative and quantitative methods with little/limited interaction between the two sources of data during the data collection stage, with the interaction happening at the data interpretation stage only in a way where findings from either method complement each other. She defines sequential triangulation as using the results of one method to inform the planning/design of the next method. She illustrates the combination and use of the two methods by suggesting a notation system. She illustrates this system with a (+) sign to denote the simultaneous collection of quantitative and qualitative data, an arrow(→) to designate that one form of data collection follows another, upper case letters indicating the method with the dominant emphasis (e.g. QUAL, QUAN) and lower case letters indicating less emphasis (e.g. qual, quan).

The discussion around triangulation illustrate a gradual shift in the conception of its purpose from that of simply methods and construct validity to broader issues of research comprehensiveness and later proposals for structured approaches in its design and implementation. It is noteworthy that the use of multiple methods for research comprehensiveness has the underlying intent of integrating the multiple findings.

This period was also typified with studies that integrated fieldwork and survey methods with the rationale that quantitative methods could make important contributions to fieldwork and vice versa. In this regard, we discuss the work of Sieber (1973) and Brewer and hunter (1989) in the following sections.

Sieber's (1973) study provides a statement of how fieldwork and survey methods can be integrated. He identifies the following contributions of each method to the three distinct stages of either research process. For a survey, he observes that fieldwork can contribute to the development of a meaningful survey design by assisting in the formulation of the research problem (or questions), the development of hypotheses, and the identification of suitable respondents. At the data collection stage, he argues that exploratory interviews and observations can yield valuable information about the

receptivity, frames of reference and spans of respondent attention in addition to helping in developing quantitative instruments. At the analysis stage, he notes that fieldwork can assist in the analysis and interpretation of survey data by: providing a theoretical structure; validating or providing plausibility for the survey findings; clarifying puzzling responses; assisting in the interpretation of results; and providing illustrative case studies. He adds that surveys can provide statistical profiles of the population within which the study is to be conducted, enhancing the fieldwork design. At the data collection stage, he observes that a survey of the population characteristics can help to ensure that the selection of the informants in the fieldwork is not biased.

With reference to what they term the 'four major research methods', i.e. fieldwork (natural observation), surveys, non-reactive research (unobtrusive observational techniques), and experiments, Brewer and Hunter (1989) argue that while individual research methods may be flawed, the flaws in each are not identical. They therefore advocate for a multi-method approach, which deliberately combines different types of methods within the same investigation, as a strategy for overcoming each method's weaknesses and limitations. They argue that using two or more research methods in an investigation provides the opportunity for cross-validation and cross-fertilization. They contend that when the findings from multi-method research converge, researchers might accept these findings with far greater confidence than if a single research method's findings are the sole source of interpretation. Acknowledging that multi-method research is nothing new, they argue for a planned, systematic synthesis of different research methods with the purpose of improving social science knowledge. They explain that multi-method research could be used for theorizing and theory testing, problem formulation and data collection, sampling and generalization, hypothesis testing and causal analysis, social problem and policy analysis, and the writing and publication of results.

These discussions on the use of multiple methodologies also underscore the intent of integrating the qualitative and quantitative components of the study. The first illustrates how the methods can be used to enrich each other for rigorous research while the second advocates for such integration.

A synthesis of the discussions of these researchers reveals an emphasis on two issues: - the integration of methodologies and the integration of methods within a study guided by a single paradigm. It also illustrates a shift over the period from somewhat technical/methodical discussions (how to mix methods within a specific approach – usually one dominant approach) to broader issues that transcend simply combining methods and techniques. Initially focused on issues of corroboration of findings and reliability of methods used, the notion of triangulation is gradually broadened to include purposes of complementarity, completeness, inclusiveness and comprehensiveness. It is apparent that the intent of integrating the multiple methods and methodologies used is central to the realization of the proposed purposes.

## 2.4. Philosophical paradigms undergirding mixed methods research

### 2.4.1. The paradigm debate

According to Howe (1988), the paradigm debate was founded on the argument that "Positivist and interpretivist paradigms underlie quantitative and qualitative methods respectively; the two kinds of paradigms are incompatible; therefore, the two kinds of methods are incompatible". This argument was premised on the writings of Kuhn (1970) who coined the phrase "the incompatibility of paradigms", which we briefly elaborate upon in the following paragraph to put this discussion in context.

Kuhn (1970) developed his argument premising on two notions: -- "Normal science" and "Paradigms". He conceived "Normal science," as meaning "…research firmly based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice". He referred to a "paradigm" as "some accepted examples of actual scientific practice…", including shared consensus and commitment to the same rules and standards for scientific practice (Phillips, 1975). He argued that paradigms are prerequisites for normal science, and that there is simply no way of practicing science without some notion of the undergirding paradigm. He contended that "scientific revolutions" are an indication of a change in paradigm ("paradigm shift"), and that the former and the emerging paradigms are incompatible. He states:

> *"The normal-scientific tradition that emerges from a scientific revolution is not only incompatible but often actually in- commensurable with that which has gone before."*

Phillips (1975) elaborates on this "incommensurability" in his interpretation,

> *"…So with regard to scientific theories, it is impossible for someone simultaneously to understand two theories formulated from within different paradigms…"*

Advocates of the 'incompatibility thesis' criticized the combination of qualitative and quantitative methods on grounds that the alleged compatibility is merely apparent, noting, "the idea of the possibility of combining quantitative and qualitative methods ignores deep epistemological difficulties and ultimately rests on the epistemologically suspect criterion of "what works" (Howe, 2003). Contrasts of the ontological and epistemological assumptions of the philosophical paradigms (Positivism and interpretivism) underlying either research methodology were the basis for this argument. Teddlie and Tashakkori (2003) cite advocates termed by Smith (1994) as 'paradigm purists' who argued that researchers who try to combine the two methods are doomed to failure due to the inherent differences in the philosophies underlying them. They contended that one paradigm precludes the other and it is therefore impossible for one to talk about using both in the same study. This is clearly a result of a strong interpretation of Kuhn's (1970) notion of the "incommensurability of paradigms"

Denzin and Lincoln (1994) underscore and elucidate the philosophical contrast of qualitative and quantitative paradigms in their analysis of the five points of difference between the two paradigms. They conclude that these differences "…reflect commitments to different styles of research, different epistemologies, and different forms of representation. Each work tradition is governed by a different set of genres". Greene and Caracelli (2002) elaborate on these epistemological difficulties by elucidating on the issues fronted by advocates for the "incompatibility thesis" related to the mixing of paradigms while mixing methods. They note that these advocates challenged the sensibility of mixing paradigmatic assumptions, particularly key assumptions like (a) the nature of the social world we endeavor to understand (ontology); (b) the nature of the knowledge we can have about that world, including relations between the knower and known (epistemology); and (c) the purpose and role of social inquiry in society.

The argument proffered by the 'incompatibilitists' was premised on the impossibility of combining qualitative and quantitative methods within the 'same study'. According to Howe (2003), "the incompatibility thesis permits only 'disjunctive' combinations of quantitative and qualitative methods within the same study, in which different methods are applied to different questions but in which the study as a whole pre-supposes different epistemological paradigms". He adds that the incompatibility thesis bars 'conjunctive' combinations of methods, in which different methods may be applied to the same question and in which the study as a whole presupposes the same epistemological paradigm. Three interpretations are inferred from this clarification to summarize the issue of contention under the "incompatibility thesis": -

1. A single study can have and use both qualitative and quantitative paradigms and therefore methods, implying that multiple paradigms can be used in a single study as long as 'methodological purity' of either is maintained.
2. That a study with more than one question, each question being answered using a different methodology but with no single overall epistemological paradigm is not an issue of contention within the 'incompatibility thesis'.
3. When a study is directed by a single research paradigm, only one set of research questions and therefore methods (either quantitative or qualitative) have to be used to answer any number of questions therein.

### 2.4.2. Paradigm stances in mixed methods research

Resulting from the 'paradigm wars', a number of stances emerged regarding the role of philosophy in the mixing of quantitative and qualitative methods. These stances fall under two categories: -

- Those that contend that paradigms do not guide practical inquiry decisions and therefore are not critical.
- Those that contend that paradigms are central to research and play a guiding role in practical inquiry decisions.

The discussion below focuses on these two positions.

*Category 1: Paradigms are not central to social inquiry*

This category includes the '***A-paradigmatic***' and '***substantive theory'*** stances. The 'A-paradigmatic' stance considers paradigms as completely irrelevant to research practice, arguing that methods and paradigms are independent of one another, disregarding the

issue of the epistemology-method link. This argument was fronted by Patton who argues that: -

> *"… in real world practice, methods can be separated from the epistemology out of which they emerged". Patton (1990) cited by Teddlie & Tashakkori Abbas (2003)*

The 'substantive theory' stance views paradigms as only helping researchers to think better but not in any way directly guiding the practice.. While this stance does not literally ignore the role of paradigms, it eschews the restrictions imposed by strict adherence to epistemology. As put by Greene and Caracelli (2003),

> *"…inquiry decisions are made not for their congruence with particular sets of philosophical assumptions but rather for their ability to further the substantive agendas of the inquiries…the nature of the concepts being studied leads the inquirer's field decisions".*

Proponents of the substantive theory stance argue that an inquirer's philosophical beliefs and understandings are entangled in his/her views of the concepts they are studying. Philosophical assumptions are therefore part of a multitude of influencing factors on an inquirer. Greene (2007) refers to these as 'mental models', representing important and intertwined strands of the individual 'mental model' that guide his/her inquiry. She considers them to include: - substantive theory, disciplinary perspectives, philosophy of science, methodological traditions, education and training, contextual factors, political factors and personal values. Because philosophical issues are mired in other competing and possibly more practical and evident influences, the role of paradigms to an inquirer's decision making is not clearly defined.

*Category II: Paradigms are central to and guide social inquiry*

This category can be sub-divided into three further sub-categories: - Those that argue against the possibility of mixing of the traditional paradigms; those who argue that the traditional paradigms could be mixed; and those who propose alternative paradigms to justify the mixing.

The first category - ***the 'PURIST' stance*** (Lincoln and Guba, 1985) is against any possibility of mixing paradigms. The argument is,

*"…the assumptions of different traditional paradigms are fundamentally incommensurable. Each paradigm represents a coherent whole, which must be respected and preserved". Greene (2008).*

This is the original Kuhnian view – which he himself later rejected. Advocates of this stance conclude that mixed methods research is therefore not possible.

The second sub-category considers the use of multiple paradigms and includes the 'complementary strengths', 'dialectic' and 'a different paradigm applicable to a specific mixed methods design' stances.

***The 'complementary strengths'*** stance posits that,

*"The assumptions of different traditional paradigms are not fundamentally incompatible, rather different in important ways. These differences are valuable and should be preserved to maintain methodological integrity while expanding the scope of the study".*

Advocates of this stance argue,

*"…methods implemented within different paradigms should be kept separate from one another".* (Brewer and Hunter, 1989; Morse, 1991; Stern, 1994)

The rationale for keeping the methods separate is the threat to the validity of either method (Teddlie and Tashakkori, 2003) and the possibility of methods losing their strengths when incorporated into each other (Brewer and Hunter, 1989).

***'The Dialectic Stance***' posits that researchers can use multiple paradigms in their mixed methods study. It assumes that all paradigms have something to offer and that the use of multiple paradigms contributes to greater understanding of the phenomenon under study (Teddlie and Tashakkori, 2003). Greene (2007) elaborate on this understanding in her statement that the 'dialectic stance' seeks,

*"Understanding that is woven from strands of particularity and generality, contextual complexity and patterned regularity, inside and outside perspectives, the whole and its constituent parts, change and stability, equity and excellence, and so forth. That is it seeks not so much convergence as insight…the*

*generation of important understandings and discernments through the juxtaposition of different lenses, perspectives and stances; in a good mixed methods study, difference is constitutive and fundamentally generative".*

Greene and Caracelli (2003) call it the "juxtaposition of opposed or contradictory ideas, to interact with the tensions invoked by these contesting arguments, or engage in a play of ideas". They however admit that the ideas or forces of contention are more at the practice than the philosophical level in her statement,

*"This does not mean that the conversations or dialogues in dialectic inquiry are necessarily about philosophical assumptions; rather, these conversations are more productively about the phenomena being studied".*

They add that in addition to philosophical assumptions, crude mental models are also privileged as key influences on inquiry decisions. Proponents of this stance view the various paradigmatic assumptions, crude mental models, context and theory all offering a partial but valuable lens on human phenomena.

***A different paradigm applicable to a specific mixed methods design:*** This stance is advocated for by Creswell et. al (2003) who argue that each mixed methods design has an applicable paradigm and that researchers need to decide which paradigm is relevant. Creswell and Plano (2007) add,

*"…This perspective maintains that investigators may view mixed methods strictly as a 'method'…thus allowing researchers to employ any number of philosophical foundations for its justification and use".*

While some authors (Tashakkori and Teddlie, 2003) categorize the 'complementary strengths' and 'dialectical' stances as distinct, others (Creswell and Clark, 2007) refer to them interchangeably, varying in the 'how' of dealing with the multiple paradigms and the way this affects the interaction between the related methods.

***Alternative/a single paradigm thesis***

This stance posits that historical philosophical incommensurabilities among paradigms are reconcilable through new, emergent paradigms such as pragmatism, scientific realism, or transformation-emancipation (Greene, 2003). Pragmatism and the

transformative-emancipatory paradigms have been the most prominent among the alternative paradigms and are reviewed in the following sections. However, the primary advocate for the transformative-emancipatory paradigm introduces it as a framework for specifically addressing social justice issues within the context of mixed methods research (Mertens, 2007) and not necessarily as a general paradigm. Teddlie and Tashakkori (2003) note that it is better conceptualized as a purpose of a research project, citing alternatives like simple curiosity which are just as legitimate. Figure 2.2 below summarizes these paradigm stances.

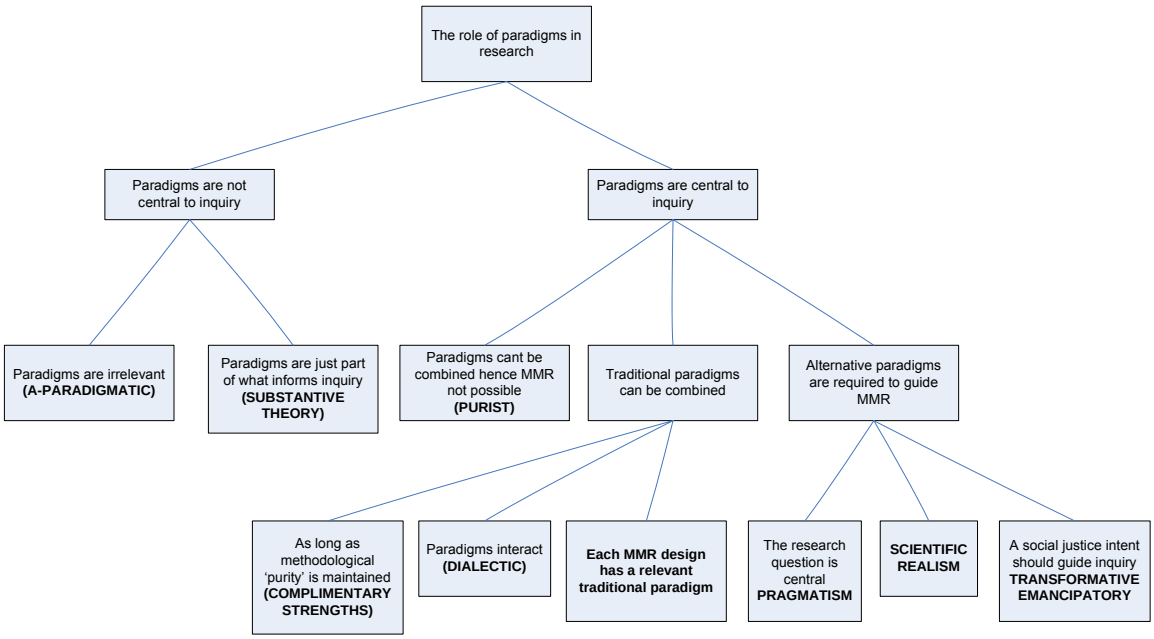


Fig. 2.2. A summary of the respective paradigm stances

Two positions towards the identification of the philosophical partner to mixed methods research stand out from this foregoing discussion. They include: -

- That which advocates for a compatibility of the traditional paradigms.
- That which challenges the 'metaphysical paradigm' and identifies alternative epistemological approaches.

The first position (dialectic and the complementary strengths stances) posits a 'compatibility thesis' in direct contrast to the arguments fronted by the incompatibilitists. They contend that multiple paradigms can be used in a single study. While the complementary strengths stance clarifies that the different paradigms used ought to be

kept separate, they are generally not very explicit about how these paradigms are used in a mixed way. Additionally, they admit that paradigms are just one of many other influences to research decisions, eschewing the central role and guidance of the multiple paradigms used.

The second position (alternative paradigm) which breaks loose from the philosophical framework defined by the 'metaphysical paradigm' and conceptualizes research in a broader context seems to have produced fewer controversies. The centrality of the research question in the pragmatic approach and the two-way mutual adjustment between the philosophical and practice realms of research appeals to both practitioners and researchers. Its flexibility in linking or fitting abstract philosophical issues to practice and vice versa makes research a reflective and growing endeavor and at the same time does not impose any a priori requirements on practitioners on what should guide their research even when they are not aware or believe in them. However, some researchers argue that it is still not clear how pragmatism guides research practice. Greene (2008) questions where the consequentialist actionable assumptions about social knowledge that are advanced in most pragmatic philosophies show up in practice. She also questions what the knowledge that integrates knowing and acting looks like and how it can be validated.

The pragmatic paradigm comprises a large component of the mixed methods 'movement' and is therefore discussed in a related section that follows.

### 2.4.3. The mixed methods 'movement'

Beyond the stances that characterized the 'paradigm wars', a 'movement' of mixed methods research emerged as a possible solution to the standoff between the traditional paradigms. Three issues characterized this 'movement': - the argument against the foundations of the incompatibility thesis, contending that quantitative and qualitative methods were not incompatible; the positioning of Pragmatism as the philosophical paradigm for MMR; and establishing mixed methods research as a new paradigm equal to the pre-dominant qualitative and qualitative approaches.

This movement can be traced from advocates of the 'compatibility thesis' who contended that methods from various paradigms could be combined in a single study. Writings of

researchers arguing from a pragmatic perspective fore-grounded this 'movement'. The early authors like Reichardt and Cook (1979) argue that that although specific research methods and techniques were sometimes linked to methodological paradigms, the paradigmatic perspective which promoted this incompatibility between the method-types was in error. They contend that evaluators and researchers often want to sample attributes from each paradigm and should therefore not be constrained to either one of the traditional paradigms when they can have the best from both. Cook (1985) coining the term '*critical multiplism'* argues that research questions can be examined from different perspectives and it was often useful to combine different methods with different biases. Howe's (1988) argument is against the metaphysical dictates of paradigms over methodology and proposes the pragmatic view that "paradigms must demonstrate their worth in terms of how they inform, and are informed by, research methods that are successfully employed". The movement was also paralleled by a gradual softening or shift in position by advocates of the 'purist' stances. Johnson. et. al. (2007) particularly cite qualitative purists (Lincoln and Guba) who started by acknowledging that quantitative data could be used in a naturalistic study (Lincoln and Guba, 1985); they later note that responsive evaluation could use whatever information (qualitative or quantitative) is responsive to the unresolved claim, concern or issue (Guba and Lincoln, 1989); later noting that "both qualitative and quantitative methods may be used appropriately with any research paradigm" (Guba and Lincoln, 1994); then reiterating that "within each paradigm, mixed methodologies (strategies) may make perfectly good sense" and clarifying that the 'argument' is not about method (Guba and Lincoln, 2005). Other writers like (Onwuegbuzie, 2000) have argued that the dichotomy between qualitative and quantitative research is false. He challenges the assumptions held by quantitative purists of objectivity, which overlooks substantive subjective decision making like the determination of the level of significance testing as 5% and the impracticality of random sampling in social research. They also challenge the relativism of extreme realists of multiple, contradictory yet valid accounts of the same phenomenon. They conclude that assumptions of both purists are self-refuting, with the positivists' verifiability principle being neither empirical nor logical while truth in the relativists only holds in the relative sense and could be false under other frameworks.

Towards furthering this movement, some researchers sought to explore the paradigm-method relationship in an effort to move the paradigm conversation beyond the stances.

Greene and Caracelli (1997) explore the relationship between inquiry paradigms and practice, and suggest that a reciprocal, mutually respectful, dialogic relationship between philosophical frameworks and methodological decisions is most warranted. They also explore the notion of the nature of paradigm attributes that matter most in mixed-method and propose a "…moving beyond the dead-end pre-occupation with age-worn, irreconcilable paradigm attributes (such as objectivity versus subjectivity) to a new analytic space". They argue, "This new space can encourage creative and imaginative mixed-method conversations, filled with multiple ways of knowing and acting-conversations that are generative and transformative in their potential insights and import".

One of the central discussions of the mixed method 'movement' was the positioning of pragmatism as its philosophical paradigm. An early advocate for this paradigm was Howe's (1988) who advocates for a philosophy of MMR premised on an argument against the quantitative-qualitative "incompatibility thesis". He advances an illustration of the differences that exist between quantitative and qualitative methods at both the practice and philosophical levels. At the practice level, he challenges the supposed incompatibility at the data, design, analysis and interpretation stages. At the data stage, he uses two senses (a 'measurement' and 'ontological') to categorize qualitative and quantitative data and defined four possible kinds of data. He illustrates that combining qualitative and quantitative data is very possible whether considering either an ontological or a measurement sense. At the design and analysis stage, he compares the mechanistic (pre-defined design, non-judgmental or objective) approach of the quantitative method with the non-mechanistic (evolving design, value-laden and subjective) ways of qualitative methods. He argues,

> *"In any study, there are only bits and pieces that can be legitimated on 'scientific' grounds. The bulk comes from common sense, from prior experience, from the logic inherent in the problem definition or the problem space. Take the view of the literature, the conceptual model, the key variables, the measures, and so forth, and you have perhaps 20% of what is really going into your study…And if you look hard at that 20%, if for example, you go back to the prior studies from which you derived many assumptions and perhaps some measures, you will find that they, too, are 20% topsoil and 80% landfill". Huberman (1987).*

Howe (ibid) argues that while quantitative methods are eminently "objective" and "scientific", the statistical inferences thereof are "as credible as their background assumptions and arguments" and that these are not amenable to mechanistic demonstration. He downplays the purity of the posited 'objectivity' in quantitative approaches, alluding to the fact that a lot of qualitative judgment goes into its analysis.

At the level of interpretation of results, Howe (ibid) contends that while either research (qualitative or quantitative) makes interpretations based on evidence from results, both are wary of alternative interpretations of their data. He argues that interpretation of results is thus at most "highly qualitative (non mechanistic) or highly quantitative (mechanistic) and that the "quantitative" "qualitative" nomenclature is a matter of emphasis. He concluded that there is no 'pure' quantitative method in real research practice.

At the epistemological level, Howe (1988) builds a case against the implied notion of the one-way conduct of research where the research paradigm dictates the methods to be used. Arguing for a "compatibilist thesis", he contends that some form of equilibrium is reached between paradigms and respective methods, alluding to the need of both concepts (paradigms and methods) in research and a two-way adjustment process. He observes,

> *"Rather than divorcing paradigms from the conduct of research (but nonetheless having them to dictate what is to count as legitimate knowledge), the compatibilist can insist on a mutual adjustment between the two such that practice is neither static and unreflective nor subject to the one-way dictates of a wholly abstract paradigm".*

Howe writes that this "compatibilism' is not a call for a reverse of the argument posited by the "incompatibility thesis", but a call for 'dialogue' in the paradigm-method relationship, where there is intimate connection and mutual adjustment between the two concepts. It is on this premise that he proffers Pragmatism, arguing that

> *"Eschewing this kind of "tyranny of method"….of the epistemology over the practical, of the conceptual over the empirical – is the hallmark of pragmatic philosophy".*

Morgan (2007) develops a case for proffering pragmatism as the philosophical basis for MMR from a broader context by reviewing the paradigm concept in social science methodology. He reviews researchers' conceptions of the meaning of a 'paradigm' and emerges with four versions representing their understanding of its meaning. First, he notes that all treat paradigms as "shared belief systems that influence the kinds of knowledge researchers seek and how they interpret the evidence they collect". The four versions he distinguishes by the level of generality of the related belief system include – 'Paradigms as worldviews', 'Paradigms as epistemological stances', 'Paradigms as shared beliefs among members of a specialty area', and 'Paradigms as model examples of research'. For the most general, i.e. "paradigms as worldviews", he points to the influence of individual worldviews on deciding what to study and how to study it. He contends that this has little to do with helping to understand the need for combining methods and how to go about it. Similarly for the version of "Paradigms as epistemological stances", he contends that while this has been the most popular and draws attention to the deeper assumptions made in research, it too tells little about more substantive decisions such as what to study and how to do so. He advocates for the third – "paradigms as shared beliefs among members of a specialty area", arguing that while it has received less attention than the former, may have more relevance. He cites the shifting in beliefs among social researchers in reference to the rise in the legitimacy of qualitative research and work in mixed methods which is largely emphasizing pragmatism. He includes the fourth – Paradigms as exemplars for "reasons of completeness".

Based on Kuhn's (1970, 1974) work on paradigms, Morgan (2007) further develops his argument by reviewing how the 'metaphysical paradigm' *(as a belief system of thinking about methodological issues in the social sciences),* premised on the 'epistemological stance', is used for comparing and defining paradigms through a trilogy of the concepts of ontology, epistemology and methodology. In challenging the 'metaphysical paradigm' and its emphasis on the "incommensurability" of defined paradigms, he notes three anomalies therein: - how to define paradigms; whether those paradigms are incommensurate; and the extent to which metaphysical assumptions actually guide research in the social sciences. He therefore envisages a "political" or "social-movement-based" account of who gets to define and draw boundaries around paradigms. He contends that this perspective views paradigms as ongoing struggles

between competing interest groups and consequently the issue of incommensurability as an emphasis of the processes of communication and persuasion about the actual work within a specialty area. He concludes,

> *"This shift from a view of paradigms as enduring epistemological stances to dynamic system of belief within a community of scholars calls into question the meta-physical paradigm's basic attempt to "impose order" on the practices in social Science research through an externally defined, a priori system from the philosophy of knowledge".*

It is against this background that Morgan (ibid) proffers the "Pragmatic approach" as the alternative to the "metaphysical paradigm". First, he presents the alternatives pragmatism has to the anomalies noted. With regard to defining paradigms, he argues that focus should be placed on issues like "line of action", "warranted assertions" and "workability" as opposed to ontological assumptions. In comparing it to the metaphysical paradigm, he argues that it means "giving up on the assumptions that there is some external system that will explain our beliefs to us". Regarding the issue of incommensurability of paradigms, he observes that "the pragmatic approach denies an a priori basis for meaningful communication between researchers" and that emphasis is placed on "shared meanings and joint action instead". For the third anomaly, research questions take pre-dominance in guiding research rather than metaphysical assumptions.

Morgan also considers the retaining of valuable contributions from the metaphysical paradigm for the development of the pragmatic approach. Particularly he focuses on the importance of the linkage between epistemology and methodology and the place of worldviews in the work of researchers. Contrary to the top-down privileging of ontological assumptions in the metaphysical paradigm, he conceives methodology as being central, connecting issues at the abstract level of epistemology and the mechanical level of actual methods.

Later writings that characterize this 'movement' question the need to differentiate or define research along the qualitative and quantitative lines. Writers like Schwandt (2000, 2006, cited by Johnson, et .al, 2007) take a stronger position on the "paradigm wars," contending that it is highly questionable whether a distinction between qualitative inquiry and quantitative inquiry is any longer meaningful for understanding the purpose and

means of human inquiry. They argue that all research is interpretive and that a multiplicity of methods is suitable for different kinds of understanding, and that it is no longer very useful for a researcher to align oneself with a particular set of methods.

Another category of more recent writers has explored whether and to what extent MMR is developed as a distinct inquiry approach. Using a framework of four interlocking but distinct domains that she notes 'characterize' a typical research approach, Greene (2006, 2008) explores the developments in mixed methods research that would justify its standing as a distinct social inquiry approach. For the first domain – 'Philosophical assumptions and stances', she writes, "A social science methodology is importantly rooted in issues that are the substance of the philosophy of science, in particular….ontology and epistemology". Citing the various mixed methods paradigm stances and their respective guidance for practical inquiry decisions, she identifies a number of issues in this domain that warrant continued attention, particularly pointing out two especially important ones. The first relates to what actually influences inquirers' methodological decisions in practice, referring to evidence from a few empirical studies, which revealed that paradigms are never cited as important practical influences. The second relates to the 'heavily favored' alternative paradigm stance, which she says is not explicit on exactly just how it is intended to influence researchers' methodological decisions. She refers to the second domain – 'Inquiry logics' as 'methodology' in social science which she writes, "…identifies appropriate inquiry purposes and questions, broad inquiry strategies and designs, sampling preferences and logic, criteria for quality for both methodology and inference, and defensible forms of writing and reporting". She adds, "…a strong inquiry logic is substantiated by coherence and connection among the constituent parts". Greene (2007) notes that progress has been made on most components that constitute a social science methodology including – purpose, design, sampling, methods choice, analysis, quality criteria, writing up and reporting. She however believes that the design dimension – 'methods characteristics' widely heralded in the early days as the rationale for mixing has not been at the forefront of mixed methods development and needs further attention. Greene (2008) also refers to the little conceptual or empirical work on how to choose particular methods for a given inquiry purpose and mixed methods purpose and design. She also raises the issue of "around what does the mixing happen?", proffering a mixing at the level of construct or variable. She defines the third domain – "guidelines for practice", as offering specific

guidelines for inquiry practice, i.e. particular steps and procedures in methods of data gathering, analysis, interpretation and reporting. This domain also "…concerns how to ask good inquiry questions, construct a defensible design from among the options available, and conduct appropriate sampling…analyses…" She surmises that substantive work is needed in this domain, proposing that this work could productively concentrate first on identifying the unique aspects of mixed methods practice that deal specifically with mixing. She defines the fourth domain – "social political commitments" as the "location of the inquiry in society", with a focus on identifying propriety roles for social science in society and providing value-based rationales and meanings or the practice of social inquiry. According to her, issues related to this domain have been on the "mixed methods radar screen" for some time particularly with reference to contributions of Mertens (2003) and others who have argued for the use of mixed methods in service of transformative and emancipator intents. She however argues that this domain remains unsettled, a position she views as a good thing, representing the grand potential of mixed methods social inquiry.

In summary, the mixed methods 'movement' was premised on the pragmatic argument that it is inevitable or common to use or combine both qualitative and quantitative methods in a single study, not to mention that is was being done without any problem. This pragmatic perspective was the basis for challenging the QUAL/QUAN dichotomy and the foundations or interpretations of the "incompatibility thesis". While some researchers disregarded the role of paradigms in research to justify the mixing of methods, others underscored that research practice has to be reflective if it is to progress, proposing a number of paradigm options that would justify the approach.

### 2.5. Definitions of mixed methods research

A number of definitions for MMR have emerged over the past two decades, clearly reflecting the varied thinking of researchers about what this research approach is. One central issue in the literature has been a discussion about what is being mixed in MMR. Creswell and Tashakkori (2007) examine various perspectives that mixed methods scholars have taken when discussing and writing about this topic. They identify four different (but not necessarily mutually exclusive) perspectives, which we would like to use in categorizing the various definitions that have been proposed. They define the first - a methods perspective, which views mixed methods primarily as a method focus on

developing and using strategies for collecting, analyzing, and interpreting multiple types of quantitative and qualitative data. The second - a methodology perspective, views mixed methods as a distinct methodology that "integrates aspects of the process of research such as worldview, questions, methods, and inferences or conclusions". The third - a paradigm perspective, focuses on an overarching worldview or several worldviews that provide a philosophical foundation for mixed methods research. The fourth - a practice perspective considers mixed methods research as a means or set of procedures to use for combining research designs, whether these designs are survey research, ethnography, or others. These perspectives are not necessarily mutually exclusive but elicit two distinct meta-categories: - a methods/practice category that emphasizes techniques, methods and procedures of research and a methodology/paradigm category which in addition includes aspects of worldviews to research.

Johnson et. al (2007) gather definitions from a number of leaders in MMR about how they define the term. This results in nineteen responses, for which they carry out a thematic analysis and draw some conclusions about an emerging definition. Building on this list of definitions, we identify others, which we categorize in table 2.2 below under the two categories defined above. Another categorization of the methods is whether they are explicit about the integration/combination of the methods (i.e. 'mix') or not (i.e. 'use). A tick (√) is used to identify each definition in this regard.

**Table 2.2 Categorization of mixed methods definitions**

| A methods or practice perspective | | Mix | Use |
|---|---|---|---|
| 1. | **Jennifer C Greene; Valerie J. Caracelli; Wendy F. Grahamn (1989)**<br>Those designs that include at least one quantitative method (designed to collect numbers) and one qualitative method (designed to collect words), where neither type of method is inherently linked to any particular inquiry paradigm | | √ |
| 2. | **John W Creswell, Vicki Plano Clark, Michelle L. Gutman, William E. Hanson(2003)**<br>A mixed methods study involves the collection or analysis of both quantitative and /or qualitative data in a single study… | | √ |
| 3. | ***Anthony J Onwuegbuzie and Nancy L Leech (2006)***<br>Involves collecting, analyzing and interpreting quantitative and qualitative data in a single study or in a series of studies that investigate the same underlying phenomenon. | | √ |
| 4. | **Pat Bazeley (**Johnson et. al, 2007)**:** involves the use of more than one approach to or method of design, data collection or data analysis within a single program of study, with integration of the different approaches or methods occurring during the program of study, and not just at its concluding point. | √ | |

| | | | |
|---|---|---|---|
| 5. | **Valerie Caracelli (**Johnson et. al, 2007)**:** a study that planfully juxtaposes or combines methods of different types (qualitative and quantitative) | √ | |
| 6. | **John Creswell (**Johnson et. al, 2007) **:** a research design (or methodology) in which the researcher collects, analyzes, and mixes (integrates or connects) both quantitative and qualitative data in a single study or a multiphase program of inquiry. | √ | |
| 7. | **Huey Chen (**Johnson et. al, 2007)**:** a systematic integration of quantitative and qualitative methods in a single study | √ | |
| 8. | **Steve Currall (**Johnson et. al, 2007)**:** involves the sequential or simultaneous use of both qualitative and quantitative data collection and/or data analysis techniques. | | √ |
| 9. | **Al Hunter (**Johnson et. al, 2007)**:** Mixed methods is a term that is usually used to designate combining qualitative and quantitative research methods in the same research project. | √ | |
| 10. | **Marvin Formosa (**Johnson et. al, 2007)**:** the utilization of two or more different methods to meet the aims of a research project as best as one can. | | √ |
| 11. | **Steven Miller (**Johnson et. al, 2007)**:** a form of evolving methodological inquiry… which attempts to combine in some logical order the differing techniques and procedures of quantitative, qualitative and historical approaches. | √ | |
| 12. | **Donna Mertens (**Johnson et. al, 2007)**:** …the use of qualitative and quantitative methods that allow for the collection of data… | | √ |
| 13. | **Udo Kelle:** the combination of different qualitative and quantitative methods of data collection and data analysis in one empirical research project. | √ | |
| 14. | **Michael Q. Patton (**Johnson et. al, 2007)**:** inquiring into a question using different data sources and design elements ... | | √ |
| 15. | **Isadore Newman (**Johnson et. al, 2007)**:** a set of procedures that should be used when integrating qualitative and quantitative procedures … | √ | |
| 16. | **Hallie Preskill (**Johnson et. al, 2007)**:** the use of data collection methods that collect both quantitative and qualitative data. | | √ |
| **Methodology and/or paradigm perspective** | | | |
| 17. | **Abbas Tashakkori and Charles Teddlie (**Johnson et. al, 2007)**:** a type of research design in which QUAL and QUAN approaches are used in type of questions, research methods, data collection and analysis procedures, or in inferences. | | √ |
| 18. | **Janice Morse (**Johnson et. al, 2007)**:** a plan for a scientifically rigorous research process comprised of a qualitative or quantitative *core component* that directs the theoretical drive, with qualitative or quantitative *supplementary component(s)*. | | √ |
| 19. | ***John W Creswell, Vicki Plano Clark, (2007:)*** a research design with philosophical assumptions as well as methods of inquiry. As a methodology, it involves philosophical assumptions that guide the direction of the collection and analysis of data and the mixture of qualitative and quantitative approaches in many phases in the research process. As a method, it focuses on collecting, analyzing, and mixing both quantitative and qualitative data in a single study or series of studies. | √ | |
| 20. | ***Johnson and Onwuegbuzie (2004); Johnson, Onwuegbuzie, and Turner (2005):*** Mixed research is formally defined here as the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods , approaches, concepts or language in a single study or set of related studies. | √ | |
| 21. | **Margarete Sandelowski (**Johnson et. al, 2007)**:** at the technique level as the combination of, e.g., purposeful & probability sampling, open-ended and closed-ended data collection techniques, and narrative and muti-variable analyses—i.e., in which anything can be used together (linked or assimilated into each other)—or …at a larger theoretical/paradigmatic level as using divergent approaches to inquiry together. | √ | |
| 22. | **R. Burke Johnson, Anthony J. Onwuegbuzie and Lisa A. Turner (2007)**: the type of research in | √ | |

| | | | |
|---|---|---|---|
| | which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) | | |
| 23. | ***Jennifer Greene*** **(**Johnson et. al, 2007)**:** Mixed method inquiry is an approach to investigating the social world that ideally involves more than one methodological tradition and thus more than one way of knowing, along with more than one kind of technique for gathering, analyzing, and representing human phenomena, all for the purpose of better understanding. | | √ |
| 24. | **Lyn Shulha (**Johnson et. al, 2007)**:** the purposeful application of a multiple person, multiple perspective approach to questions of research and evaluation. | | √ |

It emerges from this table that the majority of definitions have a methods/practice perspective. This perspective is also embraced within most definitions under the methodology/paradigm category. Creswell and Tashakkori (2007) observe that under the methods/practice perspective, researchers can use any paradigms they want to, because quantitative and qualitative methods are not "inherently linked to any particular inquiry paradigm'' (Greene et al., 1989). This perspective is labeled ''quasi-mixed'' by Teddlie and Tashakkori (2006), criticizing it for separating methods from paradigms and worldviews. It is also interesting that half of the definitions specifically mention *'integrating/combining/mixing'* the qualitative and quantitative methods, with the other half only mentioning *'the use of'* both methods/approaches, pointing to two perspectives with regard to the understanding of mixed methods among researchers. The first seems a broad understanding where mixed methods research involves multiple methods without the explicit intention of combining the methods, while the second is a stricter perspective that requires that the methods/methodologies be integrated or combined at some point in the study.

### 2.6. Mixed methods research questions

According to Tashakkori and Creswell (2007), a strong mixed methods study starts with a strong mixed methods research question or objective, which demand or require the use and integration of both qualitative and quantitative approaches or methods. They explore the issue of framing of a research question in a mixed methods study through a review of some models and mixed methods practices and observe three possibilities: -

i. Mixed methods studies need at least one explicitly formulated mixed methods question or objective following the qualitative and quantitative types of questions and emerging from the specific need to use mixed methods.

ii. Mixed methods studies will benefit from at least one overarching mixed (integrated, hybrid) question that provides the possibility of subsequent qualitative and quantitative types of sub-questions.

iii. Mixed methods studies benefit from a dynamic process in which the component (strand) questions are re-examined and reframed as the two or more strands of the study progress.

Onwuegbuzie and Leech (2006) also describe the nature of MMR questions as they explore the linkage between research questions and MMR data analysis. Noting that mixed methods researchers make use of the pragmatic method and system of philosophy, they underscore the centrality of the research question in guiding the MMR design through all the stages of a study. They compare the qualitative and quantitative kind of research questions, noting that quantitative research questions, unlike their qualitative counterparts, tend to be very specific in nature, falling into one of three categories: descriptive, comparative, and relationship. They add, "Good quantitative questions should identify the population and dependent variable(s), whether they represent descriptive, comparative, or relationship research questions. If they represent comparative or relationship research questions, then the independent variable(s) also should be identifiable". They view qualitative research questions as "open-ended, evolving, and non-directional", tending to seek, to discover, to explore a process, or describe experiences. They conceive mixed methods research questions as embedding both a quantitative research question and a qualitative research question within the same study. They add, "…a mixed methods research question necessitates that both quantitative data and qualitative data be collected and analyzed either concurrently, sequentially, or iteratively before the question is addressed". From a review of mixed methods research questions for different quantitative and qualitative designs, they note that for the purposes of mixed methods studies, "the quantitative and qualitative research questions are most aligned or compatible with respect to underlying paradigm and methods used when both questions are open-ended and non-directional in nature, and they both seek to discover, explore, or describe a particular participant(s), setting, context, location, event, incident, activity, experience, process, and/or document".

It is apparent from this discussion that both sets of authors emphasize that within mixed methods research, the framing of the questions has to dictate an integration of the

qualitative and quantitative methods that may be used. While the former set of authors focus on dictating this integration through structuring of the qualitative and quantitative question, the latter seems to imply that a mixed methods question is more congruent with a qualitative question, emphasizing an exploratory inquiry approach.

### 2.7. Mixed methods research designs and typologies

Mixed methods research designs and typologies have dominated the discussions within mixed methods research. Numerous designs which some researchers have noted are 'overwhelming' have been proposed together with a few typologies. In this section, we start with a review of four proposed typologies to tease out advance organizers/criteria that seem to inform the design of a mixed methods research. We verify the prominence of these criteria against some designs that have been proposed.

Greene et. al (1989) propose a conceptual framework for mixed methods evaluation designs. Their study is based on a review of four theoretical starting points (triangulation, multiplicism, mixing methods & paradigms, and mixed methods design strategies) towards defining their mixed methods conceptual framework. Using a review guide with seven components of the framework (purpose, design characteristics, utilization, data analysis, contexts, management and resources), they carried out a comprehensive review of fifty-seven mixed-methods evaluation studies to relate to the results from the theoretical review. They identify five mixed methods purposes. They define the first (*triangulation*) as seeking convergence, corroboration, correspondence of results from the different methods. They note that a design with a '*complementarity'* purpose seeks to measure overlapping but also different facets of a phenomenon, yielding an enriched elaborated understanding of that phenomenon. The '*development'* purpose seeks to use the results from one method to help develop or inform the other method, with a rationale of increasing the validity of constructs and inquiry results. The '*initiation'* purpose seeks the discovery of paradox and fresh perspectives and has a rationale of increasing the breadth and depth of inquiry results and interpretations. The '*expansion'* intent seeks to extend the breadth and range of inquiry with a rationale of increasing the scope of inquiry through the selection of the methods most appropriate for multiple inquiry components. These purposes are determined through the configuration of seven design characteristics which the study generates. These form the advance organizers and include: - methods, phenomena, paradigms, status, implementation: independence,

implementation: timing and study), each represented as a continuum of two contrasting attributes. They define the '*methods'* characteristic as representing the degree to which the qualitative and quantitative methods selected for a given study are similar to of different from one another in form, assumptions and strengths. They refer to '*Phenomena'* as the degree to which the qualitative and quantitative methods are intended to assess totally different phenomena or exactly the same phenomenon. The '*Paradigm'* characteristic refers to the degree to which the different methods types are implemented within the same or different paradigms. 'Status' represents the degree to which a study's qualitative and quantitative methods have equally important or central roles vis-à-vis the study's overall objective. 'Implementation: Independence' refers to the degree to which the qualitative and quantitative methods are conceptualized, designed and implemented interactively or independently. 'Implementation: timing' refers to the sequential or concurrent conduct of the methods. The 'study' characteristic refers to either a single or multiple studies within which the mixed methods design is implemented.

Creswell et. al (2003) with reference to the core design assumptions identified by Morgan (1998), Tashakkori and Teddlie (1998) and Greene and Caracelli (1997) propose four criteria (Implementation, Priority, Integration, and theoretical perspectives) for guiding a mixed methods research design. The first ('implementation') refers to the sequence of collecting both quantitative and qualitative data. The options include gathering data concurrently or sequentially, with the sequence relating to the objectives sought. They write that the implementation decision calls for clear identification of the core reasons for collecting both quantitative and qualitative data, and understanding the interrelationship between the different phases in the data collection. The 'priority' criterion relates with the pre-dominance of either method (quantitative or qualitative) in the research. They note that the researcher could emphasize one over the other, or both could be equal. They however add that this decision is difficult and problematic in practice. 'Integration' defines the stage(s) in the research process within which the mixing occurs. They identify four stages of a typical research where this could happen (research questions, data collection, data analysis and interpretation). The 'theoretical perspective' criterion has to do with the use of a theoretical lens in MMR. They particularly refer to the transformative ideology (Greene and Caracelli, 1997) whose aim is at promoting change.

Collins, et. al (2006) differentiate between the rationale for the study/approaches (i.e., why the study is needed) and the research/mixing purpose (i.e., what will be undertaken in the study) towards developing a model that incorporates the two aspects in mixing quantitative and qualitative methods. From a review of 494 articles, they develop a model that combines the reasons (i.e. rationale) used by mixed-methods researchers to combine quantitative and qualitative research and the specific purposes to which the method combination is used. From a content analysis of the articles, they emerge with four themes (rationales) comprising a total of 65 descriptors (purposes) grouped under each. The rationales identified include: - participant enrichment, instrument fidelity, treatment integrity, and significance enhancement. Participant enrichment represents the mixing of quantitative and qualitative techniques for the rationale of optimizing the sample. Instrument fidelity refers to steps taken by the researcher to maximize the appropriateness and/or utility of the instruments used in the study. Treatment integrity represents the mixing of quantitative and qualitative techniques for the rationale of assessing the fidelity of interventions, treatments, or programs. Significance enhancement focuses on enhancing researchers' interpretations of data. They also introduce three more dimensions to the model: - the first focusing at the stage (before, during or after) of the study the rationale/purpose is applicable; another on the sequencing of the methods (QUAN and QUAL); and another on the priority of either method (QUAL and QUAN).

From a content analysis of the various MMR designs, Leech and Onwuegbuzie (2007) propose a 3-dimensional typology. They use three dimensions to define this typology viz: - level of mixing (partially mixed versus fully mixed); time orientation (concurrent versus sequential); and emphasis of approaches (equal versus dominant status). They propose a 2 (partially mixed versus fully mixed) X 2 (concurrent versus sequential) X 2 (equally versus dominant status) matrix, yielding the following eight categories. (a) partially mixed concurrent equal status designs; (b) partially mixed concurrent dominant status designs; (c) partially mixed sequential equal status designs; (d) partially mixed sequential dominant status designs; (e) fully mixed concurrent equal status designs; (f) fully mixed concurrent dominant status designs; (g) fully mixed sequential equal status designs; and (h) fully mixed sequential dominant status designs. Their primary dimension is "the level of mixing" and they make a contrast between a partially mixed design which combines

results from either method only at the point of interpretation/making inferences while a fully mixed design mixes quantitative and qualitative techniques within one or more stages of the research process or across the stages of data collection, analysis and interpretation.

Two categories of criteria emerge from this discussion: - those focusing on the rationale and purpose for mixing the methods and those emphasizing the way the QUAL and QUAN methods interact within the integration. The rationales/purposes proposed by Collins, et. al (2006) seem to be dimensions of a design criteria other authors have referred to as the 'stage of integration' with participant enrichment happening at sampling; instrument fidelity at data collection, treatment integrity happening within the whole design, and significance enhancement at the stage of making inferences. It is therefore proposed that the five purposes (triangulation, complimentarity, initiation, development, expansion) proposed by Greene, et. al (1989) stand unique and different from the design characteristics.

With reference to the design characteristics proposed by Greene et. al (1989), it is argued that the 'study' criterion is not consequential as a single study is presumed in most conceptions of mixed methods research. In the same vein, it is assumed that the methods being mixed in a study are from alternative paradigms and we do not view the 'methods' criterion as very central. Therefore, six common design criteria emerge that researchers use to classify the different mixed methods designs. These include: -

1. *Implementation*: The dependence/independence of the methods. This also applies to the sequence of implementing the QUAL and QUAN components.
2. The *Priority* given to the QUAL and QUAN components.
3. *Integration:* The stage(s) of the research at which the integration of the methods is done
4. The *theoretical perspectives/paradigms* used in the research
5. *Phenomena*: The extent to which the study focus of either method is similar or different.
6. *Paradigm:* Whether the research is being implemented within a single or multiple paradigms.

In the following sections, we describe some designs that have been proposed and explore how the five rationales and six design criteria identified apply to them.

Patton (1990) proposes four mixed form designs conceived within the framework of a single research design. The first is an experimental design in which qualitative data is collected and content analysis is done. The second is similar to the first with the difference in statistical instead of content analysis. The third is naturalistic inquiry having qualitative data collection with statistical analysis and the fourth is similar to the third, with the difference in quantitative as opposed to qualitative data. Patton's designs are based on the use of a given set of methods belonging to a given paradigm within an overall alternative paradigm of inquiry. Only the 'Paradigm' design characteristic is applicable to these designs with all the methods being implemented under a single paradigm. No mention is made regarding the rationale for the mixed methods approach.

Creswell (1995) outlines five types of mixed methods designs: (i) two-phase studies where one approach proceeds or is preceded by the other, both being independent of each other; (ii) Parallel/simultaneous studies in which the methods are used concurrently; (iii) equivalent status designs where both approaches play an equal role in the research; (iv) dominant-less-dominant, where one approach is dominant and the other plays a complementary role; and (v) mixed methodology, representing the highest degree of methodological mixing where the combining of approaches happens at many or all stages of the research process. It is evident that these designs include criteria of: - implementation, priority and integration. Again, no mention is made regarding the rationale for the designs.

Greene and Caracelli (1997) categorize mixed methods designs into 'component' and 'integrated' designs. They define 'component' as those where the methods are implemented as discrete aspects of the overall inquiry and remaining distinct throughout the inquiry. The combining of the methods happens at the interpretation/conclusion stage of the study. The component designs include triangulation, complementary and expansion designs. They define integrated designs as those that integrate elements of disparate paradigms and argue that these have the potential to produce "…significantly more insightful, even dialectically transformed understanding of the phenomenon under investigation". They identify four design types under this category: - Iterative designs

which are characterized by a dynamic and ongoing interplay over time between the different methodologies associated with different paradigms; Embedded or nested designs where one methodology is located within another; 'Holistic designs', which inform a conceptual framework of inquiry that highlights the interdependence of different methodologies for understanding complex phenomena fully; and Transformative designs which give primacy to the value-based and action-oriented dimension of different inquiry traditions. These designs include the following criteria: - Integration, sequencing (iterative versus concurrency), priority (embedding/nesting), and a theoretical perspective. They however are not explicit about the rationale for the various designs.

Tashakkori and Teddlie (1998) distinguish between mixed methods and mixed model designs. They define mixed methods designs as those that combine the qualitative and quantitative approaches into the research methodology of a single study or a multi-phased study. The five specific design types proposed as falling under this category include:- sequential studies, parallel/simultaneous studies; equivalent status designs, dominant-less dominant studies, and "designs with multi-level utilization of approaches in which researchers utilize different types of methods at different levels of data aggregation". In sequential mixed methods designs, the researcher conducts a qualitative phase of a study, and then a separate quantitative phase. In parallel/simultaneous mixed methods designs, the qualitative and quantitative data are collected at the same time and analyzed in a complementary manner. For multi-level mixed methods designs, the data from more than one level of organization or groups are used to reach more comprehensive inferences regarding behaviors and/or events. They define mixed model designs as those that combine the qualitative and quantitative approaches within different stages of the research process. Conceiving a research process as comprised of three stages (the nature of the investigation, data/operations, and analysis/inferences), they argue that while mixed methods relate to the data collection stage only, mixed models concerned all three stages of research. They develop taxonomies for both categories of designs. They classify the mixed methods designs into three: - Equivalent status designs: Sequential (QUAN/QUAL and QUAL/QUAN) and Parallel/simultaneous (QUAN+QUAL and QUAL+QUAN); Dominant-less dominant: Sequential (QUAN/qual and QUAL/quan) and Parallel/simultaneous (QUAN+qual and QUAL+quan); and Designs with multilevel use of approaches. They classify the mixed model designs into six: - Type I and II are confirmatory mixed

methods studies, using qualitative data/operations but differentiated by the analysis/inference approach (statistical versus qualitative); Type III and VI are exploratory mixed methods designs using quantitative data/operations but differentiated by analysis inference approach (statistical versus qualitative); Type V is a confirmatory investigation, using quantitative data and qualitative analysis/inferences; and type IV is a exploratory investigation, using qualitative data and statistical analysis/inferences. These designs include the following criteria: - Integration, sequencing, priority and integration. Again, there is no mention of the rationales for the respective designs.

Morgan's (1998) designs are based on two central decisions: - method priority and sequencing. The priority design decision determines the extent to which either the quantitative and qualitative method will be a principle tool for gathering data. The sequencing design decision is a determination of when the complementary method comes into play (before or after the principle). Using a framework of the two design decisions of method priority and sequencing in a 2X2 matrix, he proposes four basic designs each with a different purpose: A principally quantitative method preceded by a complementary qualitative method (qual → QUANT); A principally quantitative method proceeded by a complementary qualitative method (QUANT → qual); A principally qualitative method preceded by a quantitative method (quant→ QUAL); A principally qualitative method proceeded by a quantitative method (QUAL→ quant ). Design characteristics of sequencing and priority are evident in these designs. Additionally, the rationale of 'complimentarity' is implied.

Creswell et. al (2003) propose six designs which include: - The sequential explanatory design in which: - the collection and analysis of quantitative data is followed by the collection and analysis of qualitative data; priority is given to the quantitative data; and the two methods are integrated during the interpretation phase, with a purpose of explaining and interpreting a primarily quantitative study. The sequential exploratory design in which: - the collection and analysis of qualitative data is followed by the collection and analysis of quantitative data, priority given to the qualitative aspect of the study, with a primary focus of exploring a phenomenon. The sequential transformative design takes on either of the former designs but includes a theoretical perspective that guides the sequencing of the methods. Under the concurrent triangulation design, the quantitative and qualitative data collection are concurrent and happen during one phase of the study; both having equal priority and data integration happening during the

interpretation phase. For the 'concurrent nested' design, there is only one data collection phase during which both types of data are collected simultaneously, with one being the predominant. The method with less priority is embedded or nested in the dominant one and may be addressing a question different from that addressed by the dominant method. The data from the two methods is mixed during the analysis phase of the project, calling for a transformation of one type of the data. The concurrent transformative design includes a theoretical perspective to the former two designs and is reflected in the purpose or research questions of the study. The basis of the theoretical perspective guides the choice of whether it takes on a triangulation or nested design. These designs include design criteria of implementation, priority, integration, and a theoretical perspective. They also include the rationale of complimentarity.

In a more recent study, Tashakkori and Teddlie (2003) differentiate between mono-strand and multi-strand mixed designs. In mono-strand designs, only one research method (QUAL or QUAN) is used to answer research questions that are exploratory or confirmatory, with the mixing occurring across three stages/spheres of the research process, i.e. Sphere of concepts; experiential sphere (concrete observations and operation); and sphere of inferences. They identify two designs under this category: - A conceptually mixed design in which a confirmatory or exploratory question is answered using data collection and analysis techniques of the alternative method. The second is a mixed model design in which exploratory or confirmatory questions are answered through an analysis of transformed data. The justification for these designs as 'mixed designs' is that they switch approach in the method of study. The multi-strand designs have four categories (concurrent, sequential, conversion and fully integrated). The first category has two variants: - Concurrent mixed model designs in which: - there are multiple QUAN and QUAL questions, each being answered by collecting and analyzing corresponding QUAL or QUAN data; making separate inferences and integration done by combining the inferences into a meta-inference; Concurrent mixed methods designs where one kind of question is answered by collecting and analyzing both QUAN and QUAL data, and one inference made on the basis of both data sources. Under the sequential category, the second phase (strand) of the study emerges as a result of, or in response to the findings of the first phase and has two variants: - The sequential mixed methods designs which involves one type of question, two types of data that are collected in sequence and analyzed accordingly and one type of inference at the end;

The Sequential mixed model design in which questions for the second strand of the study emerge from inferences that are made in the first strand. The Multi-strand conversion mixed designs has only one method of study and one type of data. However, they emulate two strands by having both types of analysis. It has two variants: - Multi-strand conversion mixed method design where one type of question is asked and one inference is made from both analyses; and Multi-strand conversion mixed model design in which: - multiple approach questions are asked, one data type is collected, then transformed to another data type, both data analyses carried out and inferences from either data analysis are integrated into a meta-inference. The last category is the 'fully integrated mixed model' design which incorporates two or more of the previous types. It involves multiple approach questions, both QUAN and QUAL data collection and analysis, sequential and concurrent data collection, data conversion, multiple inferences and an integration of inferences to get a meta-inference. These designs embrace design characteristics of implementation and integration, and the rationale of complimenatrity.

Creswell and Clark (2007) advance four major mixed methods designs, with variants within each type. The four major types of mixed methods designs they propose include: - the triangulation design, the embedded design, the explanatory design, and the exploratory design. They provide four variants of the triangulation design: - the data transformation model in which quantitative and qualitative data on the same phenomenon are collected and analyzed separately and then the different results are converged (by comparing and contrasting the different results) during the interpretation; the validating quantitative data model, which also involves the separate collection and analysis of quantitative and qualitative data sets, but with the transformation of one data type into the other data type; the multilevel model in which both types of data are collected within one survey instrument, the qualitative items being an add-on to a pre-dominantly quantitative survey; and the multilevel model where different methods (quantitative and qualitative) are used to address different levels within a system with the findings from each level merged together into one overall interpretation. The embedded design is a mixed methods design in which one data set provides a supportive, secondary role in a study based primarily on the other data type. They discuss two variants of this design: - the embedded experimental model where qualitative data is embedded within an experimental design, used either as a one-phase or a two-phase

approach, in which the timing reflects the purpose for including the qualitative data and the qualitative data introduced before, during or after the experiment. They note that this model has also been referred to as a concurrent nested mixed methods design; and the correlational model in which qualitative data are embedded within a quantitative correlational study. The explanatory design is a two-phase mixed methods design having two variants: - The follow-up explanations model where qualitative data is used to explain or expand on quantitative results (the primary emphasis). Specific quantitative findings that need additional explanation are identified and qualitative data is from participants who can best help explain these findings. The second (participant selection model) is predominantly a qualitative study that uses quantitative data to identify and purposefully select participants for a follow-up, in-depth, qualitative study. The exploratory design is similar to the explanatory design, with the difference that the qualitative method precedes the quantitative method which is the predominant. Variants of the exploratory design include: - the instrument development model in which the qualitative phase explores the research topic with a few participants. The qualitative findings then guide the development of items and scales for a quantitative survey instrument. In the second data collection phase, the researcher implements and validates this instrument quantitatively. Under the taxonomy development model, the initial qualitative phase is conducted to identify important variables, develop a taxonomy or classification system, or develop an emergent theory, and the secondary, quantitative phase tests or studies these results in more detail. Design characteristics of implementation, priority, and integration, and rationales of complimenatrity and triangulation are evident in these designs.

It is evident from the discussion thus far that design characteristics of implementation, priority and integration are common across nearly all the proposed designs. Design characteristics of Phenomenon and paradigm are not mentioned in any of the designs. Similarly, only a few designs are explicit on a rationale for the different designs. It is evident that there is a relationship between these design characteristics (phenomenon and paradigm) and the rationales.

Two categories of designs are identified: – (i) a few that focus only on the use of multiple methods, and (ii) those that emphasize the integrating or mixing of the methods. The early designs starting with Greene, et. al (1989) embrace both aspects by including the

multiple methods perspective through the 'expansion' purpose designs. However, the later designs seem to eschew this 'multiple methods' thinking and promoted only the mixed/integrated conception of combining the qualitative and quantitative methods/results/inferences in a study.

### 2.8. Sampling in mixed methods research

Teddlie and Yu (2007) focus on how probability and purposive samples could be used to generate mixed methods samples. They define mixed methods sampling as involving the selection of units of analysis for a mixed methods study through both probability and purposive sampling strategies. They propose a typology of five mixed methods sampling strategies described below. Under the first category – 'Basic mixed methods sampling strategies', they identify two strategies which are typically discussed as types of purposive sampling techniques. The first is stratified purposive sampling (quota sampling), whose stratified nature procedure is characteristic of probability sampling, whereas the small number of cases typically generated through it is characteristic of purposive sampling. The technique has two stages, the first involving dividing the group of interest into strata and then selecting a small number of cases to study intensively within each strata based on purposive sampling techniques. The second strategy under this category is purposive random sampling and involves taking a random sample of a small number of units from a much larger target population. The second category of strategies is sequential methods sampling where the methodology and results from the first strand inform the methodology employed in the second strand. This category has two variations defined by the sequencing of the sampling methods (i.e. QUAL→QUAN or QUAN→QUAL). What differentiates this from the basic category is that the proceeding sampling procedure is informed by the research results from the previous procedure. The third category - Concurrent mixed methods sampling is used in triangulation designs where the QUAL and QUAN sampling procedures proceed concurrently and independent of each other. They identify two basic overall concurrent mixed methods sampling procedures: - (i) Concurrent mixed methods sampling in which probability sampling techniques are used to generate data for the QUAN strand and purposive sampling techniques are used to generate data for the QUAL strand, both procedures occurring independently; and (ii) Concurrent mixed methods sampling utilizing a single sample generated through the joint use of probability and purposive techniques to generate data for both the QUAN and QUAL strands of a mixed methods study. The last

strategy is multilevel mixed methods sampling, used in research examining organizations in which different units of analysis are nested within one another. For such studies, the research is focused at questions related to two or more levels or units of analysis. They conclude by noting the possibility of complex sampling strategies (not included in their typology) that involve combinations of multiple strands (QUAL, QUAN) of a research study with multiple levels of sampling within each strand. They also give eight guidelines researcher ought to consider when putting together a sampling procedure of a mixed methods study.

Onwuegbuzie and Collins (2007) propose a different typology of sampling designs in mixed methods research. Their typology is constructed from what they refer to as a mixed methods sampling process comprising of seven steps: - (a) determine the goal of the study, (b) formulate the research objective(s), (c) determine the research purpose, (d) determine the research question(s), (e) select the research design, (f) select the sampling design, and (g) select the sampling scheme. Acknowledging that there are many mixed methods designs and a number of typologies to simplify researchers' design choices, they argue that the common denominator for most mixed method designs is the timing orientation dimension as its base, i.e. the sequencing or concurrency of the QUAL and QUAN phases of the study. With reference to Greene et. al's (1989) work, they note that another decision that the researcher should make relates to the five purposes (triangulation, complementarity, initiation, development, expansion) proposed in that conceptual framework. By crossing these two dimensions, they produce a ten cells matrix indicating the research design options. Towards selecting a specific sampling design, they introduce another dimension - the relationship of the qualitative and quantitative samples which can be identical, parallel, nested, or multilevel. In an identical relationship, exactly the same sample members participate in both the qualitative and quantitative phases of the study. For a parallel relationship, the samples for the qualitative and quantitative components of the research are different but are drawn from the same population of interest. For a nested relationship the sample members selected for one phase of the study represent a subset of those participants chosen for the other facet of the investigation. A multilevel relationship involves the use of two or more sets of samples that are extracted from different levels of the study (i.e., different populations). Using the two criteria, time orientation and sample relationship, they proposed eight different types of major sampling designs that a mixed methods

researcher might use. These designs are labeled Design 1 to Design 8. Design 1 involves a concurrent design using identical samples for both qualitative and quantitative components of the study. Design 2 involves a concurrent design using parallel samples for the qualitative and quantitative components of the study. Design 3 involves a concurrent design using nested samples for the qualitative and quantitative components of the study. Design 4 involves a concurrent design using multilevel samples for the qualitative and quantitative components of the study. Design 5 involves a sequential design using identical samples for both qualitative and quantitative components of the study. Design 6 involves a sequential design using parallel samples for the qualitative and quantitative components of the study. Design 7 involves a sequential design using nested samples for the qualitative and quantitative components of the study. Design 8 involves a sequential design using multilevel samples for the qualitative and quantitative components of the study. For all these combinations, the QUAL and QUAN components take on either a higher, equal or lower priority/weight. They summarize these eight sampling options in Figure 2.3 below.

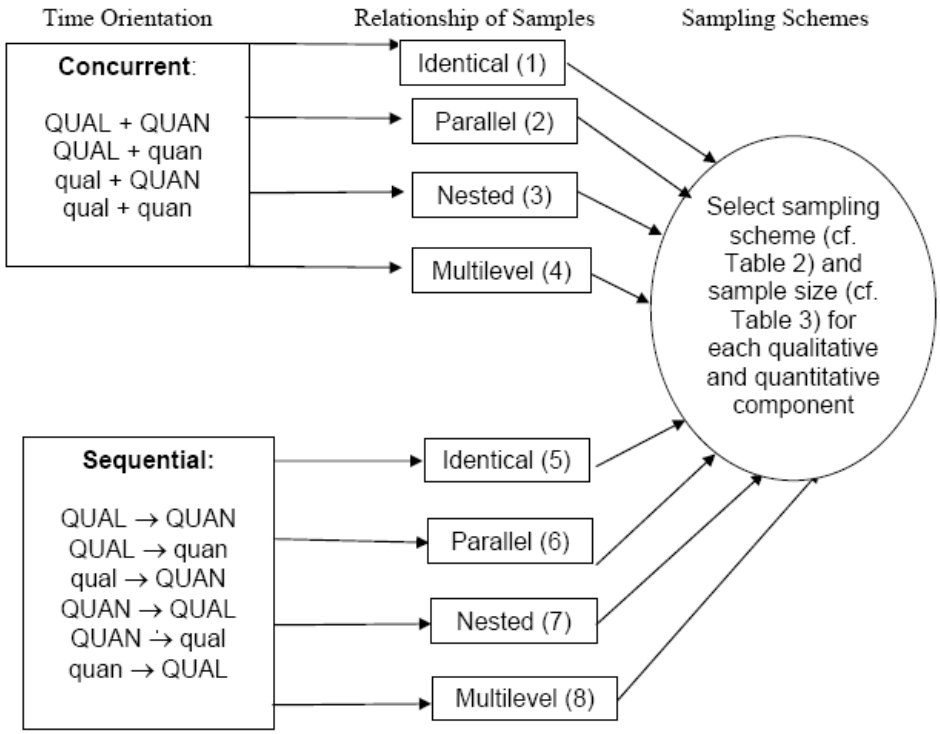


Fig. 2.3. Two-dimensional mixed methods sampling model providing a typology of mixed methods sampling designs (Onwuegbuzie and Collins, 2007)

It is apparent that Teddlie and Yu (2007) emphasize the combination of qualitative and quantitative methods to generate a single sample, consequently conceiving the mixing of methods as happening within the sampling stage. It is therefore possible within their conception that the sampling stage, the rest of the study phases can follow a single paradigm approach. On the other hand, Onwuegbuzie and Collins (2007) propose at least one sample for each methods strand in a study. A common feature among the proposed samples is that they are related to each other, requiring a mixed methods design in which either the data/methods/findings/inferences in the subsequent research phases need to be integrated.

### 2.9. Data analysis strategies in mixed methods research

Caracelli and Greene (1993) propose four integrative data analysis strategies for mixed-methods evaluation designs, and discuss and illustrate their appropriateness for different kinds of mixed-methods intents. These include: - data transformation, typology development, extreme case analysis and data consolidation/merging. This study was motivated by the empirical results of Greene et. al (1989) which they argue demonstrate that a number of authors for most studies that were reviewed did not report how they conducted their data analysis or kept both the analysis and interpretation of the data separate. The study focuses on elaborating the mixed methods analytic strategies used in those few reviewed studies which integrated qualitative and quantitative data during analysis, interpretation and reporting. The 'data transformation strategy' involves the transformation of one data type into the other to allow for statistical or thematic analysis of both data types together. In the 'typology development strategy', the analysis of one data type considers the homogeneity within and heterogeneity between sub groupings of data on some dimension of interest, yielding a set of substantive categories or typology. This typology is then incorporated into the analysis of the contrasting data type. The 'extreme case analysis strategy' involves the identification and further analysis of extreme cases which are identified through the preliminary analysis of one data type. The 'data consolidation/merging strategy' involves the joint use of both data types to create new or consolidated variables or data sets for further analysis and is "…especially suitable for mixed-methods designs with initiation intents". They suggest that these integrated analysis strategies are most appropriate when methods are mixed for the purposes of initiation. They add that the strategies are also appropriate for the expansion and development purposes, though this is not clear cut for the complementarity purpose.

However, they argue that integrated analysis strategies are generally less useful for triangulation.

Onwuegbuzie and Teddlie (2003) conceptualize seven stages of a typical data analysis process within a mixed methods framework. They include (a) data reduction, (b) data display, (c) data transformation, (d) data correlation, (e) data consolidation, (f) data comparison, and (g) data integration. They conceive '*data reduction'* as involving reducing the dimensionality of the qualitative data. *Data display* involves describing pictorially the qualitative data. This is followed (optionally) by the *data transformation* stage, wherein quantitative data are converted into narrative data that can be analyzed qualitatively and/or qualitative data are converted into numerical codes that can be represented statistically. *Data correlation* involves quantitative data being correlated with qualitized data or qualitative data being correlated with quantitized data. This is followed by *data consolidation*, where both quantitative and qualitative data are combined to create new or consolidated variables or data sets. D*ata comparison* involves comparing data from the qualitative and quantitative data sources. *Data integration* is the final stage, whereby both quantitative and qualitative data are integrated into either a coherent whole or two separate sets (i.e., qualitative and quantitative) of coherent wholes.

Creswell and Clark (2007) propose two types of analysis – concurrent and sequential. They note that these types depend on the type of MM design used which in turn is dependent on the MMR questions being asked. The concurrent data analysis approach relates to the 'Triangulation' and 'Embedded' designs while the sequential relate to the 'Explanatory', 'Exploratory' and 'Embedded' designs. They propose the following procedure for the concurrent design: - conduct separate qualitative and quantitative analysis; merge the two data sets either through data transformation or data comparison. Data transformation involves the converting of one form of data into the other form (e.g. qualitative into quantitative or vice versa) so that it is easily merged. Data comparison involves the comparing the data without transforming it through a matrix for discussion. An alternative to using a matrix is to make comparisons by examining the similarities of the quantitative and qualitative data results in the discussion section of a study. They define three stages of the sequential analysis approach - analysis of one set of data; use the results of the first stage to make decisions of what

information is most useful to stage 3; then the collection and analysis of data for the second dataset.

From a review of the foregoing discussion, two categories of mixed methods data analysis are evident: - the first involves the use of only one method (QUAL or QUAN) and requires the transformation of one data type into the alternative. This data transformation is what qualifies the integration of the methods. The second involves both analysis methods (i.e. QUAL and QUAN), which may happen concurrently or sequentially, with the integration conceived as: - the comparison of the results emerging from the respective analyses for concurrent implementation or, results of a previous method informing the conduct of the follow up method in a sequential implementation.

### 2.10. Validity issues in mixed methods research

Tashakkori and Teddlie (2003) argue that the ultimate advantage of using mixed methods is the quality of inferences made at the end of a series of phases/strands of a study. They therefore differentiate between the "results" of a study and the "inferences" that are made from that study to develop the meaning of inferential quality in mixed methods research. They note that the term 'inference' is common to both the deductive and inductive approaches to making conclusions in quantitative and qualitative methodologies and can therefore be used alike in either approach. They propose the term "inference quality" to represent the understanding of validity (internal validity for QUAN and credibility for QUAL). They define it as "…the accuracy with which we have drawn both our inductively and our deductively derived conclusions from a study". They also introduce two associated terms related to rigor in the application of method and rigor in interpretation : - the first, 'design quality' comprises the standards for the evaluation of the methodological rigor of the mixed methods research; and the second, interpretive rigor, comprises the standards for the evaluation of the accuracy or authenticity of the conclusions. They also propose the term "inference transferability" as an umbrella term to embrace the intent of generalizability or transferability as understood in the QUAN and QUAL approaches respectively. They propose four specific types of transferability: - Ecological transferability (referring to contexts other than the ones studied); Population transferability (referring to other individuals/groups or entities from those studied); Temporal transferability (different time periods); and Operational

transferability (referring to other modes /methods of measuring/observing the variables/behaviors).

Onwuegbuzie (2006) introduce the 'neutral' term legitimation in favor of the term validity in mixed methods research. While agreeing with Teddlie and Tashakkori's (2003) foregoing argument, they observe that their conceptualization of inference quality while appealing and useful needs some elaboration and extension in two ways. First, they view legitimation as a process, not just an outcome, requiring for legitimation checks at each stage of the mixed research process. Secondly, they argue that giving inference quality primary emphasis could give the false impression that other steps of the research process do not need to be scrutinized. They therefore propose nine legitimation types viz: - Sample Integration (the extent to which the relationship between the quantitative and qualitative sampling designs yields quality meta-inferences); Inside-Outside (the extent to which the researcher accurately presents and appropriately utilizes the insider's view and the observer's views for purposes such as description and explanation); Weakness Minimization (the extent to which the weakness from one approach is compensated by the strengths from the other approach); 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); Conversion (the extent to which the quantitizing or qualitizing yields quality meta-inferences); 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); Commensurability (the extent to which the meta-inferences made reflect a mixed worldview based on the cognitive process of Gestalt switching and integration); 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); Political (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). Onwuegbuzie and Johnson's (2006) framework therefore aims at verifying the quality of the mixed methods research at all stages of the research process with the use of multiple validities for the different stages of the quantitative and qualitative strands.

Dellinger and Nancy (2007) propose a framework for guiding discourse about validity in mixed methods research. Basing on the idea of construct validation as the all-encompassing purpose of having standards for any element of a study, they build on Messick's (1995) conception of construct validity to provide a unifying theme for their framework. Contending that Messick's (1009) definition of construct validity is too narrow when including qualitative data from the qualitative research traditions, they propose a re-writing it to read,

> *"…an overall evaluative judgment of the extent to which empirical evidence and/or theoretical rationale support the adequacy and appropriateness of interpretations and actions on the basis of data generated through any means".*

With regard to the notion of 'construct validity' specifically for mixed methods research, they begin by contrasting its conception for either qualitative or quantitative research. In qualitative research, they observe that constructs are negotiated by the researcher and participants through the meanings of the data. They write,

> *"Qualitative researchers' understanding of a construct emerges from immersion into the data. In qualitative research, the construct is not limited but developed and refined in a flexible, organic way".*

According to them, quantitative research on the other hand starts with a supposition that a specific construct or entity exists, basing its definition on past research, theory or phenomena. It therefore involves explicit a priori definition of the construct together with measurement processes for generating scores.

They add that the different ways by which the two approaches measure constructs does not preclude the fact that they cannot measure or are not measuring the same constructs or meanings "on the basis of systematic arrangements of perceptions, experiences, attitudes and so on". They premise their conceptualization of validity on negotiation of data meaning, basing it on the argument that "meaning is not a function of the type of data collected but rather results from the interpretation of that data". They therefore propose construct validation as a continuous process of negotiation meaning, accomplished through "argument as dialogue, criticism and objection". They argue,

> *"…construct validation is an open, continuous system in which construct meaning is the product of convergent and divergent evidence, results, consequences, and*

*arguments from all research related to that construct, whether qualitative or quantitative".*

Their validation framework uses the traditional concepts of validity from both qualitative and quantitative traditions, some new ideas proposed for mixed methods research, particularly those from Onwuegbuzie and Johnson (2006), and also introduces four new elements not discussed by earlier approaches: –a foundational element, Inferential consistency, Utilization/historical and Consequential. Using the framework involves a recursive back and forth process of negotiating meaning as guided by the validity methodologies of the various.

These discussions about validity within mixed methods research seem a build up from an initially "narrow" focus on only the inferences made to a holistic approach that focuses on all aspects of a mixed methods research process. The approach includes the traditional QUAL and QUAN validity measures (applicable where methodological purity is maintained) and novel approaches that focus on those stages where the integration of the methods, data, inferences, etc happens.

### 2.11. Criticisms of mixed methods research

Giddings (2006) cautions that while the pragmatic assumptions (having the best of both worlds, better understanding, more certainty of results, etc) of MMR sound quite reasonable on the surface, it is likely to become 'Positivism dressed in drag' if it is not open to philosophical critique. She argues that the current promotion of mixed methods is not a methodological movement but a pragmatic research approach that fits most comfortably with a postpositivist epistemology. Her first argument is with the inconsistency in the use of the terms 'methods' and 'methodology' which she notes is not only between papers but within them. She is concerned that the term used ought to communicate the researcher's thinking of what is being mixed, with a 'methods' terminology focusing on mixing the techniques of research while a 'methodology' would emphasize a mixing of a researcher's inquiry frame. Her second argument is premised on the notion of the 'thinking' of research, which she argues, "…is shaped by past experiences, epiphanies, embodied understanding and many complex and nuanced insights as well as the nature of the phenomenon to be studied". She contends that the 'thinking' of positivism continues in the thinking of mixed methods arguing,

*"The thinking in mixed-methods research rarely reflects a constructionist or subjectivist view of the world. The majority of studies use the analytic and prescriptive style of positivism, albeit with a post-positivist flavor. …A design is set in place, a protocol followed. In the main, the questions are descriptive, traditional positivist research language is used with a dusting of words from other paradigms, and the designs come up with structured descriptive results. Integration is at a descriptive level. A qualitative aspect of the study is often 'fitted in'. The thinking is clearly positivist and pragmatic".*

She surmises, "Clothed in a semblance of inclusiveness, mixed methods could serve as a cover for the continuing hegemony of positivism, and maintain the marginalization of non-positivist research methodologies".

Giddings' (ibid) criticism falls within the stance termed "***A different paradigm applicable to a specific mixed methods design"*** in which mixed methods is viewed strictly as a method and researchers are free to identify a 'best' philosophical partner to a mixed methods design at hand. It is evident that post-positivism was the pre-dominant and applicable paradigm to her study cases and those she refers to. While her concerns may be argued to be limited to the cases she studies and should not be generalized, they are worth exploring.

Symonds and Gorard (2009) challenges the categorization of the concept "mixed methods research" as a separate paradigm. Analyzing the construct and content validity, and propensity for bias, they elicited what they term "underlying problems with the concept's descriptive accuracy and exhaustiveness". From a construct validity perspective, they criticize the foundational assumption in MMR that all singular elements of research can be categorized into qualitative and quantitative paradigms, and that this is connected to practice by the assertion that strong links exist between the data types, methods and methodologies in each category. Starting with the categorizations of the data types and methods, they argued that the objectivity/close-ended versus the subjectivity/open-ended labels to quantitative and qualitative respectively and their assignment to separate paradigms is based only on assumptions about how the data is used or '*the most common use of the methods'*, as opposed to their potential or actual use. They also argue against the dichotomy of small and large sample sizes for

qualitative and quantitative respectively, noting that the quality of generalization is not restricted to any particular sample size, not to a specific research paradigm. Similarly, they challenge the categorization - 'numbers' and 'words', arguing "…in many cases, numerical data began as word, visual or audio data…" and that "…there is no justification for numbers to have a separate paradigm in the first place". Similarly, for data analysis, they argue that no method is fixed to any one paradigm, noting that the different methods can be used on either data. They conclude that as there is no strict placement of data or methods into the qualitative or quantitative domains, there is no evidence for direct normative links between methods, methodologies and paradigms.

> *"We could use the word 'quantitative' to refer only to the activity of quantification, and 'qualitative' to describe that which is examined in depth – without being linked to a research paradigm…Mixed methods itself could refer to any study which purposefully integrates multiple techniques to achieve a final set of data…".*

Symonds and Gorards (ibid) base their argument against a mixed methods paradigm on a technical/practice perspective which challenges the QUAL/QUAN dichotomy. Their criticism seems to resonate with the arguments within the a-paradigmatic stance.

It is apparent from this discussion that the criticisms against mixed methods research have been targeted at its philosophical premising and emerge from a pragmatic/practice perspective. Within this perspective, mixed methods research is conceived as a 'technical' endeavor limited within the methods/technical stages of a research process. Some critics challenge the QUAL/QUAN dichotomy (the fundamental notion in mixed methods research), while others argue that the mixing of methods is possible and has always happened within the traditional paradigms, precluding its conception as a new paradigm.

### 2.12. Analysis and synthesis of the developments in mixed methods research

This review of the literature elicits two major themes that have dominated discussions among mixed methods researchers- a methods/practice perspective and a methodology/paradigm perspective. Most discussions have focused on the methods/practice perspective, which has seemingly had least or no controversy. The majority of definitions of mixed methods research given by the researchers are oriented

towards this perspective, viewing mixed methods research as a mixing or use of multiple techniques and data. The perspective has both a 'loose' or 'weak' conception of mixed methods research where the inclusion/use of multiple methods in a single study suffices, and a 'stricter' or 'stronger' one in which the integration/combination of the methods used has to be done. The early 'classic' mixed methods studies in which the qualitative and quantitative methods co-existed under a single inquiry paradigm illustrate this stricter definition. It is noteworthy that the early multiple methods approaches broadened the conception of 'triangulation', gradually shifting the emphasis from a merely technical/methodological perspective (methods and construct validity) to one of richness of inferences resulting from an integration of multiple methods. The pragmatic orientation of the mixed methods 'movement' seems to have paved way for and expanded the options for a structured approach to designing mixed methods research with its promotion of a 'compatibility thesis'. By eschewing the dictate of paradigm over methods, it possibly facilitated the proliferation of a number of mixed methods designs that followed.

The early mixed methods designs starting with Greene, et. al (1989) abstracted the broadened notion of 'triangulation' into specific purposes, with the term triangulation limited to only the stricter version of corroboration. This brought on board a repertoire of other purposes that included complementarity, exploration, explanation, expansion, development and initiation. These purposes seem to have largely framed the thinking around the various mixed methods designs that emerged. The design characteristics of implementation, priority and integration were the most prominent within this framing. It is however noteworthy that the majority of subsequent designs placed more emphasis on the design characteristics, giving less attention to the rationales/purposes as guides for the integration of the methods.

It is further noted that while the initial purposes for mixed methods research may have embraced multiple methods research, the later designs, mixed methods research questions, sampling and data analysis approaches eschewed the multiple methods conception, emphasizing the integration of the QUAL and QUAN components. These further developments in the structuring of mixed methods research underscored the need to integrate the QUAL and QUAN components at least at one point in the research process. They therefore seem to have excluded rationales/purposes of 'expansion',

inclusion, completeness and others focusing on comprehensiveness of findings as part of mixed methods research.

This methods/practice perspective also relates to the a-paradigmatic philosophical stance, which contends that methods are independent of paradigms. It is also congruent with the foundational argument of proponents of pragmatism, which promoted a 'compatibility thesis' particularly at the methods level. It is apparent that this perspective is not necessarily a new thing within 'good' research practice. In any typical research, the choice of techniques, method and data seems to be driven by a pragmatic requirement of obtaining relevant data that can answer the research question(s) at hand. What is apparently new under the mixed methods 'movement' is the eliciting of a systematic way of carrying out research in which both qualitative and qualitative data and methods are used and integrated. The various mixed methods research designs, sampling schemes, data collection approaches, analysis techniques and validity measures that have been proposed provide a framework that can guide a researcher. Similarly, the proposed links between research purposes and designs inform a conscious and reflective approach to the rationale for the use of the various techniques. This methods/practice perspective avails researchers with numerous approaches to using multiple/mixed methods. It provides a broader conception of design when compared to the early 'classic' studies, which were largely limited to qualitative data collection within an experiment and the multiple methods studies which were focused on triangulation of methodologies.

Fewer researchers have discussed the methodology/paradigm perspective, though it has evoked the most controversy and criticism. This controversy can be traced from its premising on resolving the paradigm wars which to-date, while quieter do not seem to have resulted in a consensus. Additionally, the proponents of the various potential philosophical partners to mixed methods research are yet to demonstrate exactly how the proposed philosophical assumptions guide the practice. Some like those promoting 'pragmatism' have argued that that paradigms while important do not dictate over methods and that it is a two-way negotiation process. Others, such as proponents of the 'substantive' philosophical stance argue that what guides a researcher is a mix of paradigms assumptions and pragmatic/personal issues. This thinking varies from the traditional 'meta-physical' dictate of paradigm over method and is not explicit on the

guiding role of a mixed methods research paradigm in an apparently highly ‘practical’ endeavor. Some empirical studies that have revealed that paradigms play a very small or no role at all in guiding a researcher’s inquiry decisions underscore this. The critics of mixed methods research have also challenged the rationale for introducing a ‘new’ paradigm when the mixing of methods and data has been demonstrated and is possible/not refused under the traditional paradigms. It is apparent that the methodology/paradigm perspective of mixed methods research, while seemingly the novel notion in mixed methods research is yet to be developed to a consensual and practical level among researchers and practitioners.

# Chapter 3 – Research design and methodology

The background literature in Chapter 1 elucidates various concepts of evaluation research, particularly what the notion of valuing and the role of research methods in this endeavor mean. Chapter 2 has clarified various aspects of mixed methods research and expounded the peculiarity of this research methodology and how it ought to be carried out. Together, these two chapters elaborate and clarify the notion of mixed methods evaluation, and to a great extent guide the inquiry framework for the empirical component of this thesis.

In this chapter, various aspects of the research design and the methodology used to answer the research questions are presented. First, the research question initially developed in Chapter 1 is refined as informed by the review of the literature in Chapter 2. A description of the analytical framework informed by the literature review is presented next and is used as a basis for studying the unit(s) of analysis. A description of the population and cases for the study and the sampling methods used follows. Next follows a discussion of the data collection and analysis procedures employed. The rationale for the data analysis procedures and how it is carried out is described and finally the limitations and possible sources of error in the design and the methodology used.

## 3.1. Research questions

From the review of the literature, it emerges that the final evaluation inferences are influenced by various research components, right from the research questions through to the integration of the respective qualitative and quantitative inferences that comprise a study. The specific research questions therefore seek to explore how the different aspects of the research approach adopted contribute to the making of richer evaluation inferences. For this thesis, a richer inference is understood as either a qualitative, quantitative or combined inference resulting from or informed by an integration of at least one qualitative and one quantitative method, result, or inference. The four research questions are summarized below: -

1. What justifications (if any) for a mixed methods research approach are given by evaluators?
2. How do evaluators define and use the qualitative and quantitative methods within the context of a mixed methods research approach?
3. What is the nature of evaluation inferences made by evaluators when they foreground a mixed methods approach?

4. Are there any relationships between the rationales given and the uses of the qualitative and quantitative methods that could explain the nature of inferences made?

## 3.2. Analytical framework

The analytical framework that is expounded here is informed by the various mixed methods research dimensions discussed in Chapter 2. The literature review has revealed a number of research aspects that potentially guide the eventual methodology (inquiry approach) adopted. The following research aspects are considered: - the research aims/objectives/questions, the rationale given for using a mixed methods approach, the different methodologies and methods used, the evaluation focus of the respective methods, the nature of data collected, analysis carried out and results generated, and the synthesis of the qualitative and quantitative. These aspects are discussed in the following sections to elaborate on the analytical framework.

### 3.2.1. The research objectives/aims/questions

The pragmatic philosophical paradigm places emphasis on the centrality of the research question(s)/aim(s) in informing and guiding the eventual inquiry approach adopted. Within the context of mixed methods research, Tashakkori and Creswell (2007) and Onwuegbuzie and Leech (2006) give some insight regarding the need for 'mixed' research objectives/questions in dictating an inquiry approach that will necessarily adopt a mixed methods design. Questions of this nature dictate an inquiry approach that needs to use and integrate multiple methodologies, methods, or data or a combination of these. Our analysis of the cases therefore includes an assessment of the research questions/aims of the case studies. It specifically seeks to establish the nature of the questions, whether they prescribe a particular kind of inquiry approach and data. The objective is to determine whether mixed methods studies begin their inquiry using this a priori planning that requires the integration of the methods used.

### 3.2.2. Rationales for a mixed methods approach

A primary justification for a mixed methods evaluation approach is its potential to generate an enriched and enhanced understanding of an evaluand through explanation, elaboration, illustration, etc. This thesis seeks to establish the reasons (if any) evaluators give for using a mixed methods approach. The purpose is to clarify whether a mixed methods approach is justified as most appropriate in responding to the research questions/aims defined or whether researchers merely characterize their design as a mixed methods approach without an explicit

justification. Such rationales are inferred from the statement(s) of aims/objectives/questions or the actual uses of the different methods/methodologies included in the study.

### 3.2.3. The uses of the various qualitative and quantitative methods

Numerous mixed methods designs and typologies corresponding to different purposes have been proposed as is evidenced from the literature review in the previous chapter. This thesis sets out to establish which of these designs are common in mixed methods evaluations. Specifically, it aims to establish what the assessment focus of the qualitative and quantitative methods and data are, and how they are used in making final evaluative inferences. It particularly looks for the way the methods are integrated/combined. The study also seeks to establish whether the methods/data used are congruent with the rationales given within a study.

### 3.2.4. The nature of data, analysis and results

Our review specifically seeks to establish the nature of data that is collected, the analysis carried out and the results/inferences obtained in each study. Mixed methods research seeks to integrate various aspects of the research including questions, sampling, data collection, analysis and inferences. This specific aspect of the analysis seeks to delineate the respective research components to determine where the mixing (if at all) in the research is done, and specifically to subsequently trace how the results of the respective study components are utilized in making of the final inferences.

### 3.2.5. The synthesis of the findings from the qualitative and quantitative components

One of the primary purposes of adopting a mixed methods research approach is the generation of richer inferences from an integration of findings/inferences from the qualitative and quantitative study components. This thesis therefore places emphasis on the way the different cases synthesize the findings/inferences from the various qualitative and quantitative methods used. It specifically seeks to establish how researchers use the different inferences to avail a richer conclusion about the performance of the evaluand.

## 3.3. Unit of analysis

The Unit of analysis for our empirical study is a journal article whose study focus aims at evaluating a programme or intervention. The assessment aims at describing specific aspects of an individual study and identifying any possible relationships or patterns among such aspects towards the generation of evaluative inferences.

## 3.4. Description of the case studies

The case studies comprise fourteen evaluation research articles selected following a convenience purposeful selection approach. In selecting the studies, the study seeks to identify evaluation studies that specifically have the following attributes: - a purpose/aim of assessing the performance of a program/intervention; inclusion of a detailed description of the research methods used; use at least one qualitative and one quantitative method; evidence of qualitative and quantitative data and an indication of the data analysis carried out; provide results from the analysis of either method; and include some discussions where the inferences or conclusions could be inferred. The studies are identified through a search for articles whose titles include words or phrases like: - “mixed methods evaluation”, “combining/integrating quantitative and qualitative methods for evaluation”, or “triangulation”. The second is an assessment of the study’s background literature to verify that either the authors have an intention of integrating the qualitative and quantitative methods or classify the methods used thus. This means that the selection comprises articles that not only include qualitative and quantitative methods but fore grounded a mixed/integrated evaluation approach. In this sense, the selection is purposive as it is expected that the authors give much consideration to the issues of combining the methods within an integrated approach.

It has not always been easy to determine what qualifies a study as including a combination of qualitative and quantitative methods. This is especially with cases where the only instance of a qualitative component is the quantification of qualitative data from an open-ended question in a quantitative instrument. Others include purely qualitative studies where the only instance of quantitative methods us in the selection process. However, because the main purpose of this study is to describe, all such cases are included as long as the authors refer to the approach as ‘mixed methods evaluation’.

The selection also follows a convenience approach with the studies selected based on their accessibility to the author. The selection initially considered evaluation reports considering that they would represent more of the practice than published evaluation studies. However, it has been noted that most evaluation reports do not give details on or discuss the methods (sampling, data collection and analysis) used, and do not present the results in a logically coherent way. It is therefore not easy to tease out the study issues of interest. This potential data source has been subsequently ruled out.

The eventual selection of fourteen cases is detailed in Table 3.1.

Table 3.1. The case studies

| | Title | Publication details | Study objectives |
|---|---|---|---|
| 1. | Considerations in the Design of a Mixed-Method Cluster Evaluation of a Community Programme for 'At-Risk' Young People. | Lucke, J. C., Donald, M., Dower, J., Raphael, B. (2001). Evaluation, 7, 1, 110-113. | A comprehensive evaluation |
| 2. | Integrating Quantitative and Qualitative Methods to Assess the Impact of Child Survival Programmes in Developing Countries: The Case of a Programme Evaluation in Ceara, Northeast Brazil. | Lindsay, A. C. (2002). Health education behavior, 29, 5, 570-584. | To assess the impact of the programme. |
| 3. | Can evaluation studies benefit from triangulation? A case study. | Ammenwerth, E., Iller, C., Mansmann, U. (2002). International journal of medical informatics, 70, 237-248. | To systematically evaluate the preconditions and consequences of the intervention |
| 4. | Evaluation of a teacher mentoring program using a mixed methods approach. | Louis, T., Isadore, N., Carole, N., James, S., David, L., David, N. (2002). Annual conference for the eastern Educational Research Association. Sarosota, Florida. February 28, 2002 | Assessing effectiveness |
| 5. | Identifying Best Practices for WISEWOMAN Programs Using a Mixed-Methods Evaluation. | Besculides, M., Zaveri, H., Farris, R., Will, J. (2006). Prev Chronic Dis, 3,1, 1-9. | A process evaluation |
| 6. | Using Mixed Methods for Evaluating an Integrative Approach to Cancer Care: A Case Study. | Brazier, A., Cooke, K., Moravan, V. (2008). Integrative Cancer Therapies, 7, 1, 5-17. | To evaluate the impact on target group with respect lifestyle, quality of life, and overall well-being |
| 7. | A Mixed-Method Evaluation of a Workforce Development Intervention for Nursing assistants in nursing homes: A Case of WIN A STEP UP. | Morgan, J. C. and Konrad, T. R. (2008). The Gerontolist, 48, 1, 71-79. | To assess the impact of the intervention on specific outcome measures. |
| 8. | A Mixed-Method Evaluation of a Workforce Development Intervention for Nursing assistants in nursing homes: A Case of WIN A STEP UP. | Morgan, J. C and Konrad, T. R. (2008). The Gerontologist, 48, 1, 71-79. | An impact assessment |
| 9. | A mixed-method evaluation of nurse-led community-based supportive cancer care. | Howell, D. M., Sussman, J., Wiernikowski, J. (2008). Support cancer care, 16, 1343-1352. | A process evaluation |
| 10. | Transition services for incarcerated | Abrams, L. S., Shannon, S. K. S., | To assess the benefits |

| | Title | Publication details | Study objectives |
|---|---|---|---|
| | youth: A mixed methods evaluation study. | Sangalang, C. (2008). Children and Youth Services Review, 30, 522-535. | and limitations the programme. |
| 11. | A mixed methods evaluation of televised health promotion advertisements targeted at older adults. | Berry, T. R., Spence, J. C., Plotnikoff, R. C., Bauman, A., McCargar, L., Witcher, C., Clark, M., Stolp, S. (2009). Evaluation and Programme Planning, 32, 278-288. | To assess the impact of the programme |
| 12. | A Mixed Methods Evaluation of the Effect of the Protect and Respect Intervention on the Condom Use and Disclosure Practices of Women Living with HIV/AIDS. | Teti, M., Bowleg, L., Cole, R., Lloyd, L., Rubinstein, S., Spencer, S., Aaron, A., Ricksecker, A., Berhance, Z., Gold, M. (2009). AIDS Behav, 11, S106–S116. | To assess the outcomes of the intervention |
| 13. | Addressing the Challenges Faced by Early Adolescents: A Mixed-Method Evaluation of the Benefits of Peer Support. | Marsh, H. W. and Ellis, L. A. (2009). American journal of community psychology, 44, 54-75. | To assess the effectiveness of a the intervention |
| 14. | Using mixed methods to evaluate the Pediatric Lead Assessment Network Education Training program (PLANET). | Polivka, B. J., Chaudry, R. V., Sharrock, T. (2009). Evaluation and the health professionals, 32, 1, 23-27. | An impact assessment |

## 3.5. Data collection and analysis

Following the analytical framework discussed in the foregoing sections, each article is read and re-read a number of times first to understand its content but specifically to identify and classify the different analytical components. The five elements of the analytical framework form the organizing frame for presenting and discussing each article.

The rationale for a mixed methods approach is teased from the study aims, especially among studies where it is explicitly defined. Specifically, the role of the complimentary method (which is typically the qualitative) gives some insight to this rationale. However in some cases where it is not explicit, the rationale is teased from the concluding discussions where the authors make reference to the methodology used and the achievements thereof.

The case studies vary in the way the research questions/aims/objectives are presented. Some formulate them as questions while others state aims/objectives of the study. For a few, the related method is mentioned along with the question/aim though for the larger component, the related method is captured from the methods section of the study.

A few of the studies mention explicitly the rationale for the mixed methods approach and even include a detailed discussion justifying the approach. However, for the majority, the rationale is

established from the purpose of the qualitative component which in most cases is introduced to play a complimentary role to the primary quantitative method.

The respective method uses are tagged to the assessment focus of each method, with particular emphasis on the programme aspect each method was used to assess. The specific programme aspects include: - design, implementation, and outcomes/impact. The respective designs for the qualitative and quantitative components are identified either from what the authors mention or inferred from the way authors describe the methodology. The study seeks to establish the following about the respective methods in each case: - the pre-dominance (if at all) of one method over the other, largely established from the way the authors introduce the respective roles of the methods; and the sequencing of the methods established from either the timeframes indicated for each data collection or by the dependence of one inquiry approach on the other, guided by the inquiry questions and findings of the respective methods. It is however noted that the classification of the sequencing as “concurrent” or “sequential” is largely determined by the dependence/independence of the respective inquiries and not necessarily the timing of the different data collections.

The data collected and the analysis and results thereof are identified for each method. For the quantitative component, descriptive and inferential statistical analyses are sought and the results/findings presented as frequencies and significance tests identified. For the qualitative component, the main focus is on the identification of themes resulting from a thematic analysis.

A primary intent of this thesis is to establish how the various case studies discuss the findings/inferences from the qualitative and quantitative components. This includes an inquiry into how a given component is informed/guided by the alternative. The study particularly seeks any discussions that relate to or are informed by at least one of the results, findings or inferences from both the qualitative and quantitative components.

The collection of the data and its analysis are conducted in a recursive iterative way. The data collected inform the formation of the initial themes, which are consequently used to review the data and ensure it is effectively categorized. This process continues throughout the process of analyzing the fourteen cases, requiring the review of completed analyses a number of times basing on emerging categories and themes. The findings on each of the respective analytical components are categorized, with the analysis seeking to have at least two case studies

demonstrating a specific category that is defined. The study finally seeks to identify possible relationships between various categories of the analytical components across the case studies.

## 3.6. Limitations of the study

This study aims at characterizing evaluation studies in which researchers have a clear intention of following a mixed methods approach. The cases are subsequently identified by the inclusion of specific terms in the title of the article. It is however noted that apart from a few, the majority of the selected cases do not actually mention such an intention either from the background literature or as reflected in the actual implementation. Many simply describe or qualify their use of the qualitative and quantitative methods as mixed methods evaluation. This is evidenced from the lack of specific mixed-research questions, rationales for mixing the methods and any mention of a mixed methods design approach followed. Therefore some classifications made about the different cases are not necessarily based on the best/ideal representation of a particular mixed methods research approach. The effort of this thesis to classify the various study designs could therefore be an "over" interpretation of the different authors' perceptions to their use of the methods.

The convenience selection approach limits the scope and variety of cases that can be included. The study aims at getting a representation of as many mixed methods research study designs as are discussed in the literature review. However, the majority of the cases follows an 'expansion' purpose design and is mainly limited to the assessment of the outcome/impact aspects of a program. This has resulted in a lack of saturation for some of the themes that emerge from the analysis, especially among cases with 'complimentarity' or 'triangulation' designs.

The use of evaluation articles only as opposed to evaluation reports potentially limits the scope of characterizations of mixed methods evaluation that can be identified. This limitation arises from the fact that published evaluation articles may typically strip away aspects of the evaluation context that is fairly evident in typical evaluation reports. Such a context can include issues like: - how the evaluation questions and the methods to use are determined, the resources (time and funds) available for the study, the availability of some information like baseline data to facilitate the adoption of specific designs, the practicality of having a sequential mixed methods research design when questions are pre-defined and budgets fixed, etc. These contextual factors can

avail explanations for the research approaches adopted by practitioners, facilitating the identification of more pragmatic recommendations which this thesis lacks.

# Chapter 4 – Results and discussion

Guided by the analytical framework developed in Chapter 3, the qualitative content and thematic analysis and results of the fourteen selected studies are presented in a chronological order (following the year of publication) in this chapter. The first part of the chapter presents case analyses of the different studies. For each study, a brief description of the intervention is given to contextualize the study. This is followed by a description and analysis of the evaluation approach used and the results thereof. The last aspect is a discussion of how (if at all) the study synthesizes the qualitative and quantitative findings/inferences in making the final conclusion. The second part of the chapter is a cross-case analysis and draws crosscutting and broader inferences from the different case analyses. Within each study, patterns between study components and the inferences made are sought and themes emerging across the studies are discussed.

## 4.1. Description and analysis of the case studies

### 4.1.1. Considerations in the Design of a Mixed-Method Cluster Evaluation of a Community Programme for 'At-Risk' Young People. Lucke, et. al (2001)

*Context and programme description*

This intervention was a community development programme for young people ‘at-risk’ of self-harming behavior. It was a suicide prevention programme implemented in four separate areas of Queensland, Australia and funded by the state government over three years. The programme intended to use a primary health care approach to empower communities to become active in the prevention of self-harming and suicidal behavior among young people aged 10-24 years. The stated goal of the programme was the reduction of suicide as the main outcome measure. It was coordinated from central government offices, with multi-disciplinary programme teams located on four different areas (1 to 4) of the state. The programme teams were intended to provide an area-wide service and undertake a range of activities, including education and training, support and referral. In particular, the programme teams were intended to initially establish Local Action Reference Groups (LARGs) and to support their functioning within the regions. LARGs were to be local community coalitions made up of local people (teachers, general practitioners, police and other community members) bringing about change for young people. This group of people would be trained by the local programme team to act as a source

of information and support for the community when dealing with young people. The programme commenced during different months between May 1995 to January 1996 for the four areas.

*Description of the study*

The authors refer to the study as a comprehensive evaluation of the programme at both state and area levels. Though this is the intention, the results of both the qualitative and quantitative components indicate an emphasis on assessing the programme design and implementation for a formative purpose. No mention is made of the programme outcomes/impact.

The study seems to have intended to use a quantitative-only design, which later integrates qualitative components in response to requests by a stakeholder group. The authors state,

> *"…although the programme was to contain both qualitative and quantitative components, there was to be heavy emphasis on quantitative data"*

They add,

> *"Interviews were conducted with young people….in response to requests from programme teams for more qualitative components in the evaluation plan".*

In the final discussion, they justify the mixed methods approach, adding,

> *"The use of multiple methods was necessary because of the many types of information required…"*

In a later section discussing the mixed methods approach, the authors write,

> *"It was clear from the outset that a mixed-methods design would be required in order to fulfill the extensive requirements of the evaluation agreement".*

The authors present a detailed discussion of the mixed methods paradigm stances and relate these to their choice of a pragmatic 'outlook' in responding to consultations with programme teams. This 'outlook' seems to explain the inclusion of a qualitative component. They state,

> *"The philosophical underpinnings for the evaluation of this cluster evaluation began from the dialectical position, but quickly engaged a heavy pragmatic outlook in response to consultations with programme teams. It was clear from the outset that a mixed methods design would be required in order to fulfill the extensive requirements of the evaluation agreement"*

The authors also note that their design follows a mixed methods 'expansion' framework, though they do not mention any specific design. They write,

> *"The design included distinct components in a predominantly expansion framework…where different methods were used for different elements of the evaluation…However, there were also elements of complementarity…where qualitative information was used to clarify and contextualize quantitative results".*

It is apparent from the above statements that the rationale for the mixed methods approach is premised on the need to collect both qualitative and quantitative data. The reference to an "expansion framework" indicates awareness about the various mixed methods typologies that have been proffered. This 'expansion' framework is congruent with a mixed methods expansion purpose design that seeks to use multiple qualitative and quantitative methods to study different aspects of a programme.

The authors refer to a mixed methods purpose of '*complimentarity'*, stating that qualitative information is used to clarify and contextualize quantitative results. It is however not apparent from the presentation of the results and subsequent discussions how or where this is realized since the quantitative results are presented and discussed independently of the qualitative results as noted in a subsequent section. It is possible that they are referring to the preliminary discussions/interviews with service providers whose aim is to gain information to inform the evaluation design.

The study comprises three interviews and two surveys. One interview targets key service providers and seeks information about their involvement and experience with the programme. The primary purpose of this interview is to gain preliminary information to guide the design of the evaluation. Another interview with young people is designed to provide recommendations about programme functioning. The third interview is with program staff and aimed at providing recommendations about future programme implementation and functioning and specific evaluation of the LARG concept. In summary, the interviews focus on evaluation planning and assessing programme design and implementation. For two interview studies, the authors do not mention how the respondents are selected while for one, it is a purposive convenience selection approach. The study does not mention or discuss the qualitative analytical methods used.

For the surveys, a mix of purposive and convenience selection of cases and to a small extent random sampling is used. One survey targets service providers and aims at providing formative evaluation data for future programme development, and providing data for the impact evaluation of the programme. It includes pre- and post-surveys of the service providers, particularly focusing on the nature of services they were giving to the young people. The second survey is a baseline targeted at young people in the four programme areas. The authors state that the results of this study do not contribute directly to the evaluation but could be used for future surveys.

From the perspective of the research data used in this evaluation study, it is apparent that it comprises three interviews and one survey, both components focusing on programme design and implementation. It is noteworthy that one of the interviews is not within the typical framework of a data collection method in research but more like the informal/preliminary discussions that inform the evaluation plan. The fact that the authors include this as one of their methods may be an illustration of a broader thinking of what qualifies as a qualitative method in research.

*Results from the study*

The interviews with the service providers do not lead to any specific results, only having a purpose of informing the evaluation plan. The authors note,

> *"The first round of interviews with key service providers served to initiate contact with many service providers and raised awareness of the programme and its aims. This was also the case in the second round of interviews".*

The interviews with the young people reveal that they had positive opinions about the programme (i.e. they had no problems with accessing the programme or getting involved and that the programme had been helpful).

The interviews with programme staff result in many recommendations for programme functioning. They particularly cite the problematic nature of the LARG concept among the weaknesses.

From the survey of the service providers, the results reveal that most had concentrated on working with schools, youth services, mental health services and community services. They add

that the survey "usefully identified barriers to provision of training to service providers…and reflected the different emphases of separate programme areas".

*Discussion of the results/inferences from the qualitative and quantitative study components*

The study does not synthesize the respective findings/inferences in any of their discussions. They discuss each method together with its results and the use thereof in separate sections of the report.

### 4.1.2. Evaluation of a teacher mentoring program using a mixed methods approach. Louis, et. al (2002)

*Context and programme description*

The Teacher Evaluation Programme (TEP) was a 3-year medium sized (8,000 students), mid-west, suburban school district's intervention designed to support new teachers in the school district. The authors neither mention the location nor timing of the programme. Implemented and administered by the District Review Boards (DRBs), it was targeted at beginning teachers, experienced teachers new to the district, and reassigned teachers as well as educational personnel such as librarians and counselors working under the educational personnel certificate for the first time. This target group was referred to as interns. The district assigned consulting/experienced teachers to orient, help develop, and evaluate interns during their first three years of employment with the district. For each intern, there was a designated administrator who worked with the consulting teacher on various issues. The designated administrator also dealt directly with the intern on some matters. A consulting teacher was directed to spend between 20 and 36 hours with his/her intern in classroom observations and conferences during the first year and 13 to 24 hours in the second and third years. A very specific timeline of activities and reports was given with specific actions to be carried out during some months of the year. Stipends were provided to the consulting teachers to prepare program documents and meet with interns, administrators and the DRB.

*Description of the study*

The evaluation study aims at assessing the effectiveness of TEP and perceptions of stakeholders. The authors do not mention a specific rationale for using a mixed methods approach but with reference to their data collection and analysis approach write,

*"The combination of these two approaches is sometimes referred to as mixed methods"*

The study starts with conversations between the district administrator responsible for the teacher mentoring evaluation plan, the DRB and the researcher. The researcher then develops concepts to stimulate discussions with a six person focus group composed of representatives from all stakeholder groups in the programme. The purpose of these discussions is to identify the issues that are meaningful and pertinent for the district at the time. Through these discussions, the evaluators come up with five evaluation questions which together with the conversations held inform the 23-item questionnaire that is eventually used as the data collection instrument. The broad evaluation questions focus on: - effectiveness of the programme, areas of improvement and perceptions of stakeholders. The last questionnaire item is an open-ended question seeking comments. The authors provide a detailed description of this preliminary process which they refer to as a qualitative component of the study. Using the five questionnaires, data is collected from three categories of respondents who include: - newly hired teachers, consulting teachers (mentors), and administrators/managers. The quantitative data is analyzed through descriptive statistics, correlations and analyses of variance. The qualitative (open-ended) responses are analyzed by initial coding, focused coding and categorization.

The authors summarize this research process in the statement,

> *"In this study we started with the qualitative question, "What should be evaluated?" which was followed by a quantitative assessment of the responses to the questionnaires. These results along with the analysis of the open-ended questions resulted in a qualitative-quantitative analysis of the mentoring-evaluation program"*

*Results from the study*

The discussion of the quantitative findings focuses on comparing the significance of the differences in responses between the three categories of respondents. For the qualitative component of the questionnaire, the authors argue that the responses provided support to the survey findings and some hints on ways to improve the mentoring-evaluation situation. They particularly cite recommendations for improving the program design.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The study does not include a discussion of the mixed methods approach. The qualitative and quantitative findings and inferences are discussed separately. The authors conclude,

*"Mixing qualitative data and analysis and quantitative methods both longitudinally in the course of the evaluation and latitudinally in the questionnaires provided a final product that was highly meaningful and useful to the stakeholders".*

This study has only one quantitative research design (a survey) in which the instrument has one open-ended question that seeks comments. It is on the basis that both qualitative and quantitative data are collected by the one instrument that the authors qualify a "mixed methods" approach. Additionally, they qualify the first process of defining and clarifying the evaluation questions and instrument items as a qualitative component of the study. Although the authors mention that responses to the open-ended questions provided support to the survey findings, they do not elaborate on this "support". The recommendations for improvement that they present are not discussed in relation to the survey findings.

### 4.1.3. Can evaluation studies benefit from triangulation? A case study. Ammenwerth, et. al (2002)

*Context and programme description*

This is a study by the Heidelberg University Medical Center aimed at systematically evaluating the preconditions and consequences of computer-based nursing process documentation, with special emphasis on acceptance and attitudes issues. The study is based on a software - PIK (''Pflegeinformations- und Kommunikationssystem'', a German acronym for ''nursing information and communication system'') which was introduced on the four wards of the University Hospital of Heidelberg, Germany. The software functionality covered the six phases of the nursing care process. The study is conducted in four different (psychiatric and somatic) departments where a nursing documentation system had been introduced in 1997. On every ward, motivated nurses were specially trained as key-users. In addition, all other health care professional team members (such as physicians) received an introduction to the computer-based documentation system on how to access the records. The study period was between August 1998 and October 2001.

*Description of the study*

The authors do not explicitly refer to this study as a 'mixed methods evaluation' per se but triangulation of the qualitative and quantitative components. Their perception of triangulation includes aspects of validation and comprehensiveness of findings. They write,

*"…triangulation has two main objectives: to confirm results with data from other sources (validation of results), and to find new data to get a more complete picture (completeness of results)…"*

The rationale for a mixed methods approach emerges from the results of the quantitative study which reveal significant attitude differences across wards. The researchers seek to better understand the reasons for the different attitudes on all wards, and especially to learn more about the differences between ward C and the other wards. Specifically, the quantitative results reveal that on ward C, the overall attitude of the computer-based system remained rather negative, even during the last round of the surveys. The qualitative study is therefore initiated to answer a question regarding the reason for this.

The objective of the quantitative study is to analyze the changes in the nurses' attitudes with regard to nursing process, computers in nursing, and nursing documentation system, following the introduction of the computer-based system. The objective of the qualitative study is to further analyze the reasons for the different attitudes on the wards. They argue that while the quantitative study exactly describes these attitudes, the qualitative study is intended to further explain those quantitative results.

For the quantitative component, the authors use a prospective intervention study with three time measurements: - approximately 3 months before introduction; approximately 3 months after introduction; approximately 9 months after introduction. The three quantitative data collection incidences span a period of three years. A purposive selection of cases is used, with the nursing management selecting four study wards falling under different departments. The selection of participants within the different wards followed a convenience sampling approach where the nurses volunteered to participate. Responses from only 40 nurses (i.e. those who completed all three questionnaires) are used. The researchers carry out a descriptive and correlational analysis of the quantitative data. The qualitative study is conducted four months later, following the completion of the analysis of the quantitative study. A purposive selection of cases from the quantitative sample is used for the qualitative component. Open-ended focus group interviews are conducted for this category of respondents. The interviews are analyzed using inductive, iterative content analysis.

*Results of the study*

The study makes the following observations from the quantitative analysis: -

A quantitative analysis of the individual items of the questionnaires revealed negative effects especially on ward C. Both on wards C and D, the nurses stated that the documentation system did not ‘‘save time’’, and that it did not ‘‘lead to a better overview on the course of patient care’’. In addition, on ward C, the nurses stated that they ‘‘felt burdened in their work’’ by the computer-based system, and that the documentation system did not ‘‘make documentation easier’’. On ward A and B, the opinions with regard to those items were more positive. Overall, the results of quantitative analysis pointed to a positive attitude toward the computer-based nursing documentation already shortly after its introduction, which significant increase on three of the four wards later on. However, on ward C, the quantitative results revealed negative reactions, showing a heavy decline in the attitude scores during the second questionnaire.

From the qualitative content analysis, the following are noted. For ward C, the nursing process had not been completely implemented before. Documentation of nursing tasks covered a 24 h/day, due to the very young patients and their great need for care. Thus, the overall amount of documentation on ward C was higher. Patient fluctuation was also highest on ward C. For each patient, a complete nursing anamnesis and nursing care plan had to be established, which was in the opinions of the nurses more time-consuming than it had been before. On ward D, the attitude toward the documentation system was high. The nurses saw benefits, especially in a more professional documentation, which would lead to a greater acknowledgement of nursing. Standardized care planning was seen to make care planning much easier, without reducing the individuality of the patient. The role of the computer-based system in communication seemed to be rather clear: it supported, but did not replace, oral communication. For wards A and B, the overall attitudes were also positive. The nurses stressed the better legibility of nursing documentation.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The authors approach the study with a “meta-methodological” framework that informs their conception of “triangulation” as meeting the objectives of validation and completeness of findings. Their background literature informs a detailed discussion of the theory of ‘triangulation’ which includes the various dimensions of data, investigator and methods described by Denzin. Their approach seems a self-conscious and reflective commitment to a mixed-methods design

or methodology, which guides their synthesis of the findings from the qualitative and quantitative components.

In discussing the results from the qualitative and quantitative components, the study focuses on two issues: - ***validation*** and ***completeness*** of results. With regard to validation of results, the authors note the congruence of results from both the qualitative and quantitative. These include: - favorable attitudes on the three wards; and problems with regard to the user satisfaction with the nursing documentation system on ward C. This study aspect seeks triangulation of results without following the 'strict' triangulation design *(i.e. variance in the samples used, the questions asked, different programme aspects, and sequencing of the qualitative and quantitative components)* as discussed in the mixed methods literature.

With regard to completeness of results, the authors note the impact of the computer-based documentation system on documentation processes and communication processes which is not sought nor detected by the questionnaire. They also find some divergent results on two issues: - reduction in effort needed for documentation by the computer-based system and the effectiveness of the training.

The authors use the findings from the qualitative component to provide an explanation for the negative attitudes on ward C as revealed by the quantitative component. They note that the previous computer experience was seen as rather low on ward C, and also the number and availability of motivated key users. Then, during the introduction of the nursing documentation system, the workload was rather high on ward C due to staff shortage which increased pressure on the nurses. Additionally, nursing documentation had previously at least partly been carried out in the patients' rooms. At the time of the study, computers were only installed in the ward office. No mobile computers were available, which according to the nurses lead to double documentation. Thus, the ward was forced to change their documentation patterns which then also affected the communication patterns within the health care professional team. Even at the time of the interviews, those changes did not seem to have been completely integrated into the daily routine.

### 4.1.4. Integrating Quantitative and Qualitative Methods to Assess the Impact of Child Survival Programmes in Developing Countries: The Case of a Programme Evaluation in Ceara, Northeast Brazil. Lindsay (2002)

*Context and programme description*

The Ceara's Health Workers' Program (P.A.C.S.) was a child survival programme, an initiative of the state government of Ceara (North East Brazil) aimed at improving the ability of community members to meet their own health needs. It was initiated in 1988 in response to one of the most severe droughts in Ceara's history. The Ceara SESA recruited a number of community members to promote health-benefiting behaviors during the disaster. The programme covered all 184 municipalities in the state and employed approximately 7,953 community health workers in the interior and 1,000 in the capital city of Fortaleza. The programme had the following activities: - every month, each of P.A.C.S's community health workers (CHWs) visited the homes of 50 to 250 assigned families to provide health and nutrition education. CHWs weighed infants and children and collected information on health indicators. Pregnant women were referred to health units for prenatal care, as were sick children in need of treatment. CHWs reported information monthly to a supervisor, who aggregated the data at the municipality level.

*Description of the study*

The study is an outcome evaluation aimed at assessing the impact of the intervention. It comprises two ecological (quantitative) and one verbal autopsy (qualitative and quantitative) studies. The objectives of the ecological studies are to address the relative contribution of various determinants to differences in infant mortality rates, and variations in prevalence of inadequate weight gain in infants and young children among municipalities. The verbal autopsy study seeks to identify and examine factors associated with infant deaths.

The background and introduction to this study presents an intention of assessing the programme impact, and complimenting this with a description of the programme processes and contextual factors to provide a better understanding of programme performance, particularly for purposes of improvement. In justifying the mixed methods approach, the author observes that the predominantly quantitative methods used for assessing the impact of such programmes is a limitation, especially with regard to providing information for decision making and improvement of programmes. She argues that qualitative methods complement the quantitative, particularly with regard to the process related issues of a programme. She therefore contends that the

evaluation of child health programmes would benefit from the use of qualitative research techniques. She concludes that for this particular evaluation, qualitative methods are used to,

> *"…strengthen our understanding of the quantitative results, including gaining a broad understanding of the causes and circumstances of infant deaths, generating local explanatory mechanisms for infant mortality, and identifying potential factors associated with infant survival amenable to interventions".*

The three studies focus on assessing different programme aspects (infant mortality, inadequate weight gain, and circumstances around an infant's death). The two quantitative studies have a correlational design while the combined/verbal autopsy study is an integrated design with a corroboration purpose. The verbal autopsy study is a 3-months field research. For this study, extreme case sampling is used basing on two criteria *(six municipalities with the highest and five with the lowest reported infant mortality rates).* The data is collected through in-depth interviews using a semi-structured format and a questionnaire conducted with mothers or child caretakers. The questions for both the qualitative and quantitative components of the combined study are similar. The three studies are conducted concurrently.

*Results from the study*

From the quantitative analysis, the author presents determinants of the variance in the various outcome measures of infant mortality, diarrhea-specific mortality, and prevalence of inadequate weight gain. These include: - exclusively breast-fed infants, up-to-date prenatal care, low household income, female illiteracy rate, inadequate water supply, urbanization, per capita gross municipality product (GMP), percentage of pregnant women with prenatal care up-to-date, and participation in growth monitoring. The results of the quantitative component of the combined study (verbal autopsy) focus on frequencies, times and locations of infant deaths. The results of the quantitative component of the verbal autopsy study are frequencies on: - location of death, examination by a doctor, hospitalization and cause of death. A thematic analysis of the interview transcripts identifies recurring themes based on the most frequent answers cited by respondents. A general result of the verbal autopsy study is that the three major groups of factors that alone or in combination appeared to contribute to most deaths are: - delays in seeking medical care on behalf of the parents, medical interventions reported as ineffective by mothers, and delays in providing medical care to children who arrived at the hospital too late in the day to be scheduled for consultation.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The author includes a detailed discussion of how the qualitative and quantitative findings are triangulated in demonstrating how the qualitative component enriched the quantitative component. The author makes the following observations: -

She mentions the different dimensions of the programme that the qualitative and quantitative methods addressed. She notes,

> *"The integration of the two methods provided different data and information, answering questions from multiple perspectives, thus enriching our knowledge of health and providing a more comprehensive understanding of complex public health issues and programs.*

She adds,

> *"In this evaluation, the quantitative method used…yielded information about infant mortality rates … and prevalence of inadequate weight gain… In contrast, the use of an "anthropological approach" yielded information about health decision–making processes and an understanding of the disease or illness experience within a specific cultural context".*

She elaborates on how the qualitative findings enriched the quantitative findings, particularly citing the mechanisms of change. She notes,

> *"For example, the quantitative methods showed a significant association between female literacy level and infant mortality rates. This finding, although important, does not provide information into the potential mechanisms through which mother's educational status influences child survival. In contrast, the findings from the qualitative methods showed that one of the mechanisms through which mother's education influences an infant's chance of survival is the choice of treatment of life-threatening illnesses.*

However, this mechanism is not included among the findings of the qualitative component which identifies three infant mortality determinants that the author does not refer to. It is possible that this 'mechanism' is identified through informal discussions as the research progresses.

The author also states that the qualitative methods are used as a follow up to the conclusions of the quantitative methods.

*Qualitative methods were used as a follow-up to the conclusions from the ecological data analysis…and supplemented the quantitative research by exploring complex phenomena in more depth and enhancing the understanding of the quantitative results.*

However, the inquiry issues of the qualitative component seem to be independent of the findings from the quantitative components. It is therefore not clear what the author means by the term "follow-up".

In demonstrating the enrichment of the quantitative findings by the qualitative component, the author writes,

*"…the findings from the qualitative methods showed that one of the mechanisms through which mothers' education influences an infant's chance of survival is the choice of treatment of life-threatening illnesses".*

It is apparent from these statements that the author's discussion is focused more on presenting the multiple findings/inferences from the different studies as opposed to their integration or synthesis.

### 4.1.5. Identifying Best Practices for WISEWOMAN Programs Using a Mixed-Methods Evaluation. Besculides, et. al (2006)

*Context and programme description*

WISEWOMAN (Well Integrated Screening and Evaluation for Women across the Nation) was a programme created by the Centre for Disease Control in Atlanta, USA in 1995 and comprised fifteen state or tribal projects. The programme was designed to build upon the National Breast and Cervical Cancer Early Detection Program (NBCCEDP) by offering: - 1) screening for risk factors associated with cardiovascular disease and 2) lifestyle intervention services to women aged 40 to 64 years who participated in NBCCEDP. WISEWOMAN participants had to be uninsured and ineligible for Medicaid. The lifestyle intervention offered through WISEWOMAN was a key service intended to modify the behaviors associated with increased risk for cardiovascular and other chronic diseases. The intervention was predicated on the notion that obesity, poor diet, physical inactivity, and tobacco use could be modified to reduce high blood pressure and elevated serum cholesterol levels at relatively low cost and with minimal risk to participants. The CDC not only required all WISEWOMAN programmes to offer a lifestyle

intervention but also encouraged them to use the national guidelines for heart-healthy eating, physical activity, and tobacco cessation in developing their interventions

*Description of the study*

This evaluation study has an overarching aim of identifying best practices in implementing the lifestyle interventions in the programme. The authors justify the mixed methods approach on the argument that all methods have weaknesses and that the typical experimental approaches used for the kind of research purpose they are pursuing inevitably had limitations. They propose the use of two or more complementary methods or a mixed methods approach as alternatives. They specifically give examples of the strengths limitations inherent in all qualitative and quantitative methods, contending that combining methods ensures that the research is objective and has rich findings.

The study design includes multiple case studies focused on the collection of qualitative data only. It comprises in-depth case studies of selected WISEWOMAN projects and of high-performing and low-performing local sites within each project. The authors state that case studies allow them to explore how and why projects and local sites used certain practices, providing insight into the relationship between program implementation and program effectiveness. They add that the practices of high-performing sites are compared with practices of low-performing sites to identify if and how they differ.

The quantitative component of the design is embedded in the selection process of sites for the study. Five projects are purposefully selected and the performance of the various sites in each project rated. This involves analysis of programme data and ranking of each site's performance. Based on the ranking, one low-performing and two high-performing sites are selected from each project, realizing a total of 15 sites for the case studies. For the case studies, the qualitative data collection consists of preliminary data collection and site visits. The data collection also includes in-person interviews with staff members who played a role in developing or delivering the lifestyle intervention at the project and site levels; a focus group with WISEWOMAN participants at each local site and structured observations. Each of the different data collection methods has a specific purpose.

The authors make the following statement which may clarify their conception of a mixed methods approach: -

*"To conduct this mixed-methods evaluation, we applied the following five steps: 1) site selection using quantitative program performance data, 2) development of a conceptual framework for guiding qualitative inquiry, 3) development and refinement of data collection instruments, 4) collection of qualitative data, and 5) analysis of qualitative data to identify best program practices".*

*Results from the study*

The study does not present the best practices but talks of completion of a toolkit where they would be packaged at a later date. It however gives an example of how it went about identifying a practice of interest.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The study does not synthesize findings/inferences from the qualitative and quantitative components.

In the conclusion, the authors state,

*"Using a mixed-methods evaluation minimized the weaknesses inherent in each method and improved the completeness and quality of data collected. A mixed-methods evaluation permits triangulation of data and is a promising strategy for identifying best practices"*

The quantitative and qualitative methods are used to address different aspects of the research process (sampling versus assessment of processes). It is therefore not so clear how they minimize each other's weaknesses especially noting that at no stage are the results from both methods synthesized. The quantitative method weaknesses are used solely within the ranking of the projects and its weaknesses would not be helped by a qualitative method which is not used for this purpose. Similarly for the case studies, only qualitative methods are used and any weaknesses therein could not be addressed by the quantitative method.

The authors mention a 'triangulation' or combination of multiple qualitative data sources and informants, making reference to a 'within methods' and data triangulations. The statement below refers,

*"We also gathered data to **triangulate** or to combine and compare responses from multiple informants and sources, to develop a consistent understanding of lifestyle intervention implementation".*

They also mention in the concluding remarks that a mixed methods permits the triangulation of data. It is however noted that with reference to this study, the triangulation of the data they may be referring to is limited within the informants (data triangulation) and qualitative methods (observations and interviews) and is not contingent on the inclusion and use of the quantitative method. This 'permit' by a mixed methods approach as stated in this study is not clear. It is therefore possible that the authors qualify this use of multiple qualitative methods as a mixed methods approach.

### 4.1.6. Using Mixed Methods for Evaluating an Integrative Approach to Cancer Care: A Case Study. Brazier, et. al (2008)

*Context and programme description*

The Integrative Approach to Cancer Care was a programme run by the Centre for Integrated Healing in Vancouver, British Columbia. It provided integrative cancer care to cancer patients and their families each year. The Centre's physicians provided adjunctive care with the patients maintaining their primary relationships with their family doctors and their specialists. A foundation of the Centre's integrative approach included its Introductory Programme (IP) and physician visits, which were part of standard care. The goal of the IP and physician visits was to provide a framework to help people explore the ways in which mind, body, and spirit could contribute to healing and to support them in creating their own integrative healing programme. The first exposure that the majority of patients had to the Centre was attendance at the IP. The IP consisted of two days of seminars and experiential sessions presented by the Centre's physicians and complementary practitioners.

*Description of the study*

This study has an overall objective of evaluating the impact of participation in the integrative cancer care programme on patients' lifestyle, quality of life, and overall well-being. All new patients starting at the Centre for Integrated Healing between May and September of 2004 were invited to join the study. Forty-six of seventy-seven new patients agreed to participate.

The rationale for a mixed methods approach is understood to be based on the need to capture a more comprehensive perspective about the programme by including outcomes and process. The authors argue that integrative approaches to health care have both outcomes and process issues and demand the use of methodologies for evaluation that can account for both of these factors. They write,

> *"Using mixed methods allows for a more comprehensive approach to evaluation by enhancing both the ability to capture adequate information about the phenomena under study and the validity of data interpretation".*

They add,

> *"…qualitative research provides a means of exploring the wholeness of the participant's experience…and clinically relevant information about individual variation…as well as bring to light meaningful and desirable changes for patients that may not be captured by quantitative instruments".*

Specifically, the qualitative component is intended as a complement to the pre-dominant quantitative study *(focused on programme effects)* by addressing the process related programme issues which the authors argued are rarely captured by the quantitative approach.

The quantitative component design is a longitudinal pre- and posttest with baseline, 6-week, and 5-month data points for quantitative measures and with qualitative data collected midway through the follow-up period to "further explore programme impacts". It is however noted that the qualitative inquiry is independent of the quantitative results. The qualitative component focuses on the process/implementation issues (i.e. experiences of participants), specifically seeking to understand participants experience with the programme. A purposive convenience selection approach is used for the quantitative component and paired t-tests are used to compare differences in outcome measures from baseline. The participants from whom qualitative data is collected are selected using a convenience approach from the quantitative study sample. An interpretive description qualitative approach is utilized to examine the experience of integrative cancer care using the constant comparative method of analysis. The qualitative and quantitative components are conducted concurrently and are independent of each other.

*Results of the study*

From the quantitative analysis, there is no statistically significant differences between baseline and post measures on the five outcome dimensions of: - Functional Assessment of Cancer Therapy–General Survey, Multidimensional Health Locus of Control Survey, Medical Outcomes Study Social Support Survey, Hospital Anxiety and Depression Scale, Herth Hope Index. From the qualitative data analysis, positive opinions from patients are obtained about how the programme helped them. These include: - empowered decision making, creating personal change (i.e. self-care and self awareness and personal growth), and active engagement.

While the intention and rationale for a mixed methods and inclusion of the qualitative component is to complement the outcomes with programme process related issues, the results of the qualitative study indicate a focus on only the outcome aspects, with no mention of the process related issues as initially planned.

*Discussion of the results/inferences from the qualitative and quantitative study component*

In the discussion of the qualitative and quantitative results, it is apparent that the authors prefer the positive qualitative findings as is evidenced by their effort to explain the null effects revealed as resulting from the weaknesses in the design of the quantitative study. They write,

> *"The reasons the quantitative findings do not detect what appears to be positive change expressed by the patients from the focus groups and interviews are unclear but may be due to a number of factors, including a sample size that was too small to assess anything but very large changes; a follow-up timeframe that was not long enough to evaluate change appropriately; or inappropriate outcomes or survey instruments".*

The findings from the qualitative component 'contradict' with the quantitative findings. The authors seem to be challenged as to which inferences should take higher priority. This results in a preference of the positive qualitative findings to the negative quantitative findings, which they explain away as design limitations of the study.

In the conclusion, they state,

> *"…the mixed-method approach to the evaluation enhanced the ability to interpret the quantitative findings and provided a broader understanding of patients' perspectives than a quantitative analysis alone".*

However, it is apparent that the authors do not necessarily interpret the findings from the quantitative component but instead base on the findings from the qualitative component to the disregard the negative quantitative findings whose design seems to have apparent design limitations. The eventual inferences from this study are therefore based only on the single qualitative approach. On the other hand, it is possible that there is no contradiction between the qualitative and quantitative findings. The qualitative themes could be proximal outcome measures and different from the more distal quantitative measures. The perspectives of the patients revealed by the qualitative study could be an indication of knowledge and skills acquired as a result of participating in the programme. On the contrary, the quantitative measures *(internal control, depression, anxiety, tangible support, hope)* might required some intervening mechanisms *(e.g. time or an enabling context)* from the time of the proximal effects to be realized. It is apparent from this discussion that for the corroboration of findings, it is important to verify that both the qualitative and quantitative components are assessing the 'same' programme issue. Otherwise, potentially relevant information can be discarded or disregarded. For this study, asking the patients generally about their experience with the programme seems to focus on slightly different outcome measures from those emphasized by the quantitative study.

### 4.1.7. A Mixed-Method Evaluation of a Workforce Development Intervention for Nursing assistants in nursing homes: A Case of WIN A STEP UP. Morgan and Konrad (2008)

*Context and programme description*

WIN A STEP UP was an 'ongoing' workforce development intervention aimed at improving the working situation of Nursing Assistants (NAs) in North Carolina's nursing homes in an effort to decrease turnover and improve quality of care. The programme was a partnership between the North Carolina department of health and human services and the University of North Carolina institute on aging. The intervention was initiated in 2002 and at the time of the evaluation study, the programme was in its sixth year of operation. It provided a 33-hr curriculum covering clinical and inter-personal skills and distributed financial incentives to participants as they proceeded through the curriculum and completed their retention commitments. A core feature of the programme was that it required commitments from each NA, the nursing home management and the WIN A STEP programme staff. The programme required that approximately 10 NAs per facility agree to attend the classes and remain employed at the facility for an agreed-upon

amount of time. The facility was also required to commit staff time to completing the programme and distribute a retention bonus or wage increase to NAs who completed the programme.

*Description of the study*

The evaluation study aims at assessing the impact of the programme on: - turnover, perceived quality of care and job quality, job performance of programme participants, coaching supervision, and diffusion of the programme to other staff. The authors are not explicit about the rationale for the use of mixed methods/design. However, they make some reference to gaps in the survey data to be addressed by the qualitative component. They note that the qualitative component was to: - explore some implementation issues, address concerns that were not captured by the quantitative component, and assess impact from the perspective of a category of stakeholders. In this regard, they write,

> *"The goal of the content analysis…was to understand (a) how individual sites varied in their implementation of the programme, (b) what events happened between baseline and follow-up that may have influenced the intervention and were not captured by the survey data, and (c) what was the perceived impact of the intervention on the facility and staff from the perspectives of the managerial informants".*

The authors refer to the design as a "mixed-methods matched control evaluation design".

The quantitative component uses a quasi-experimental approach in which the implementation team recruits and retains eight participating organizations and ten comparison nursing homes. It focuses on assessing programme impact and includes four surveys. A convenience selection approach is used to select participants based on availability and willingness to participate. There is one case of random sampling in the comparison group. There is no sampling for the organizational surveys as all sites are included in the study. The qualitative component is a field study intended to focus on programme impact, implementation and contextual aspects. It is however noted that only results relating to the impact aspects are presented from the qualitative component. Data collection is through semi-structured interviews with managerial informants conducted prior to the start of the programme, 3 months after completion, and 6 months after completion. The qualitative and quantitative components are independent of each other.

*Results of the study*

The following results are presented from the quantitative analysis: -

There is no statistically significant difference between participating and comparison groups for the measures of: - turnover, workload, interpersonal care, job quality, job satisfaction, quality of co-workers, communication or resident-focused care. Participants show significant improvement when compared to controls in measures of: - nursing care, supportive leadership scores, and perceiving greater career rewards. From the qualitative analysis, the following emerge: - the managers felt that the programme had an impact in two main areas: increased job satisfaction/morale and improved quality of care. They described the participating NAs as more confident and proud and reported that NAs felt more rewarded. Workers had stronger clinical skills and knowledge; managerial informants reported that supervisory styles, skills, communication, and techniques of coaching supervision nurse participants improved.

Though the qualitative component is intended to focus on impacts, context and implementation, it only avails results for the impact measures.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The study does not synthesize the qualitative and quantitative findings or inferences. The respective results are presented and discussed separately.

There is a contradiction in the qualitative and quantitative results with regard to outcome measures of job satisfaction and quality of care. The authors however seem to prefer the positive qualitative findings, but are non-committal in their subsequent inference statement which does not state that the programme had an impact but rather similar programmes could have impacts on measures including these. The following quote refers.

> *"Several findings indicate that programmes like this one can produce desirable effects, including positive feedback from management, improved quality of nursing care and supportive leadership, perceived financial and career rewards among participants, and suggestions for a modest improvement in NA turnover".*

### 4.1.8. A mixed-method evaluation of nurse-led community-based supportive cancer care. Howell, et. al (2008)

*Context and programme description*

The Interlink Community Cancer Nursing (ICCN) was a community-based programme of nurse-led clinical case management that provided direct support and system navigation to cancer patients and their families in the home setting for over 15 years. The authors neither mention the location nor dates of the programme, only referring to a metropolitan city with a population base of 2.5 million. It was modeled after the Macmillan nurse program in the United Kingdom, which provided expert advice and support to patients and families with cancer and health care professionals through home visiting and telephone support. Interlink nurses were specialized oncology nurses who functioned as independent practitioners providing direct supportive care services through home visits and telephone follow-up and used skills of advocacy and system navigation to mobilize a network of support to meet patient/ family needs working interdependently in collaboration with other providers and community agencies.

*Description of the study*

The study is a process evaluation aimed at obtaining an in-depth understanding of the ICCN model of care and its care processes prior to summative or outcome evaluation. Its specific aims are formulated as: - to (1) examine the operations and activities of the ICCN program including the population served in order to delineate its components, supportive care processes and a theory of change for the program including hypothesized relationships to patient/family outcomes (logic modeling); and (2) assess perceived effectiveness of the model (provider reaction) in coordinating care and meeting cancer patient/family supportive care needs of its intended target population.

The authors do not explicitly state a rationale for a mixed methods approach, only mentioning that “a mixed-method process evaluation using multiple data sources with triangulation of quantitative and qualitative data was conducted”. They use purposive selection of cases to identify key informants representative of diverse health care sectors and disciplinary perspectives who had working knowledge of this model of care. They use a number of data sources that include: - site visits and program documents; semi-structured interviews for specialized oncology nurses to examine Interlink nurse qualifications and care processes; secondary data in an ICCN clinical database to determine services provided by the program for

a complete service year; and a survey to obtain service providers' perceptions of the effectiveness of ICCN on supportive care coordination and patient outcomes. The quantitative analysis is mainly descriptive statistics (frequencies, percentages, means) summarizing nurse qualifications and characteristics of the population served. A thematic analysis is carried out for the open-ended responses on the survey data and the interviews.

*Results from the study*

The findings of the study are structured around the aims of the project. For most of the programme aspects studied, they present a description of the programme that includes descriptive statistics largely from secondary data illustrated with qualitative data from the interviews. For the first aim (Examining ICCN components (structure) and supportive care processes to delineate a theory of change for the program and hypothesized relationships to patient outcomes), the authors present the following: -

1. *Descriptive statistics about the population served by the programme*. They noted,
   *"Most patients admitted to the program had advanced disease suggesting that ICCN was not reaching its intended target population of newly diagnosed cancer patients".*

They follow this with the results from the qualitative study, noting,

*"ICCN nurses acknowledged timely referral to the program as an issue as it was perceived that many patients were referred late in the illness trajectory".*

They add,

*"Interlink nurses expressed concern that these patients were struggling with many supportive care issues that could have been addressed earlier but often their distress was undetected in the cancer treatment clinic".*

2. *Length and profile of service.* From the qualitative analysis, the study describes two major categories of the patients, particularly defining the length of stay and patient service needs. The study also presents descriptive statistics including: - categories of months of service, reasons for referrals and sources of referral.

3. *Program components and supportive care processes:* The authors develop a logic model and note that "the findings of the logic modeling process was synthesized with findings from

multiple data sources in the study into a theory of effectiveness that delineates hypothesized relationships between ICCN care processes and patient outcomes (change logic)". However, these multiple data sources are not clarified.

4. *Direct oncology and nursing practice:* The authors note that a descriptive analysis of randomly selected clinical records confirmed that Interlink nurses assessed supportive care needs. They complement this with qualitative data that clarified on the different care processes and nurse roles, adding,

   *"Qualitative data from critical incidents provided further insight into the supportive care processes and role of Interlink nurses. A core process category and four sub-categories were derived inductively from critical incident interviews*.

5. *Coordinated mobilization/linking to services:* The authors discuss the referrals as revealed from the quantitative study together with themes from the interviews to describe the care support from multiple providers and the roles of the ICCN nurses in this regard.

6. *Building system capacity:* The study presents the role of nurses based on the interview data.

For the second aim (Perceived effectiveness of the ICCN model (service provider reaction) in coordinating care and meeting supportive care needs of the target population), the authors present descriptive statistics on: - opinions about oncology nurse roles and importance of the various functions of the programme, and reasons for referrals,. They illustrate each of these with data and analyses from the qualitative components of the study.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The authors include a general discussion about nursing care with reference to the various results presented in the previous section. Possibly because the discussion and relating of the findings from the qualitative and quantitative components is done as part of the presentation of the results, they do not specifically discuss this after the results section. They only mention,

*"The mixed method approach used in the study provided a comprehensive understanding of the ICCN program and its processes of care and how it could contribute to patient and family outcomes (program logic) leading to the development of a theory-based model of effectiveness that could be tested in future trials".*

The illustrations from the qualitative data focus mainly at clarifications and explanations for the quantitative findings through descriptions that elicited the workings of the programme. They also use the various multiple data sources to develop the programme logic model. They conclude by noting that the mixed methods approach provided a comprehensive understanding of the programme processes and its contribution to programme outcomes. It is noted that the authors refer to this illustration of quantitative findings with qualitative findings generally as "triangulation of qualitative and quantitative data".

### 4.1.9. The Evaluation of Large Research Initiatives: A Participatory Integrative Mixed-Methods Approach. Marcus, et. al (2008)

*Context and programme description*

The TTURC initiative was a 5-year $70 million project funded by the NCI, the National Institute on Drug Abuse (NIDA), and the Robert Wood Johnson Foundation (RWJF) and implemented at one agency (identity not disclosed) within the U.S. federal government. The initiative provided support to multiple research centers to study new ways of combating tobacco use and nicotine addiction, and to help translate the results and implications of this work for policy makers, practitioners, and the public. Each center's research portfolio covered basic and applied research as well as research on policy-relevant issues in studies being conducted at the center. One of the primary goals of the initiative was to encourage and support trans-disciplinary research. The research supported and generated by the initiative was intended to set a new direction in how tobacco-related research should be conducted. Researcher training was a major component of the initiative and included new and established investigators with the hope of broadening their scope of expertise within tobacco and across disciplines. Specific funds were provided to the centers to help facilitate the translation of basic and applied research into policy and practice.

*Description of the study*

This study is a pilot evaluation designed to gain experience with potential evaluation methods and tools and provide an assessment of TTURC processes and implementation and a preliminary exploration of short-term and intermediate-term outcomes.

The authors do not mention any rationale for a mixed methods approach but describe it thus basing on the multiplicity of measures used,

> *"The approach taken in this pilot evaluation is aptly described as mixed-methods…because multiple qualitative and quantitative measures and analyses were incorporated into the design".*

and the multiple data sources,

> *"Many of the individual measures were themselves combinations of qualitative judgmental data and quantitative indicators".*

The authors seem to understand the term 'methods' in a broader context, including approaches like concept mapping to identify evaluation measures and logic modeling as some of the methods. The overall design is therefore a combination of multiple research methodologies. The different studies are conducted concurrently and are independent of each other. The study includes the following as the methods used for the research: - concept mapping, logic modeling, a detailed researcher survey, content analysis and systematic peer-evaluation of progress reports, bibliometric analysis and peer evaluation of publications and citations, and financial expenditures analysis. These methods are briefly described below: -

- A Concept mapping that focuses on defining and identifying outcome measures through a quantitative analysis of a brainstorm of ideas by key stakeholders.
- An outcome logic model developed by arranging the clusters of the concept map in the expected temporal order.
- A Researcher survey: A self-administered questionnaire (Researcher Form) is used to elicit the opinions and evaluative assessments of the TTURC researchers regarding the entire range of outcomes in the logic model. It consisted of 25 closed-ended and three open-ended questions. The analysis involves descriptive and inferential statistics.
- A quantitative content analysis of progress report summaries is done on each of the progress reports to assess reported results and significance of the research using 14 markers.
- Peer evaluations: Eight peer reviewers evaluate the progress report summaries for the sub-projects across the seven funded centers. The Peer Evaluation Form assesses several areas: (a) The overall progress of the subproject, (b) progress in each of the outcome areas on the logic model, and (c) the impact of the research to date on four important audiences or constituencies. A quantitative analysis of these reports on the measures is carried out.

- Bibliometric analysis: A number of index variables are constructed from publication and citation data to assess the scope and quality of publication output from the research.
- A financial analysis of all the sub-projects

*Results from the study*

The logic model is used as the key unifying device for organizing and grouping results from the multiple methods for each outcome area and enabling synthesis of the findings. Results are classified according the model into the three broad temporal stages of short-term markers, intermediate markers, and long-term markers.

**Short-Term Markers**

- *Training.* The training of students, new researchers, and staff is one of the highest rated outcome areas according to TTURC researchers. On average, they assess training good to excellent. Nearly one third of all subprojects report progress in training outcomes over time.
- *Collaboration*. The results show that researchers are collaborating across disciplines and value collaboration and trans-disciplinarity. Collaboration receives the second highest progress rating of the 13 areas rated independently by peer evaluators.
- *Trans-disciplinary integration*. The ability to conduct trans-disciplinary research is the highest rated performance marker across the centers after publication quality. Researcher attitudes about trans-disciplinary research are uniformly high and positive.
- *Financial management.* There is significant variability across centers in their ability to spend allocated funds as originally proposed. Several problems are identified including significant difficulties starting up in a timely manner, delays in funding allocations from NIH, significant budget carry-overs from year to year, and significant changes in project personnel.

**Intermediate Markers**

- The authors define the intermediate markers as including the logic model categories of Methods, Science and Models, Recognition, Publications, Communications, and Improved Interventions. In terms of peer evaluation, Methods has the highest rated progress whereas Science and Models is third highest.
- In the researcher survey results, limited progress is reported overall by the researchers themselves in the development of Science and Models and Methods. On the methods side, “good” progress is reported by researchers with respect to the development of measures. In

terms of scientific theory development, “good” progress is reported in “understanding the relationships between biological, psychosocial, and environmental factors in smoking.”

- Bibliometric analyses indicate that TTURC publications are placed in well-cited journals, and TTURC publication citation rates are statistically significantly higher than for typical articles in the same journals.
- Communications of research findings is rated on the researcher survey on average as “good” by researchers. And moderately good progress is reported on the development of interventions.

**Long-Term Markers:** The authors note that even at this early stage in the initiative, TTURC researchers report considerable impact on policies at the state and local levels and on practice with respect to tobacco control.

**Pattern-Matching:** Short-term markers (i.e., process measures) show the greatest progress over time with intermediate and longer term markers showing lower progress levels. The trends over time suggest that the TTURC initiative in general is making progress along the lines that would be expected given the logic model.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The authors group the results that fall under a specific outcome measure together, though they do not synthesize or discuss the relationship between the qualitative and quantitative results. However, they cite one example which may give some insight into their understanding of mixed methods research. They state,

> *“For instance, we gathered information about publication quality that included quantitative bibliometric data on citation rates and journal quality, and subjective ratings from center researchers and multiple peer evaluators. These results needed to be summarized and synthesized with the other data collected for other outcomes on the logic model. This could not have been accomplished without use of an integrative mixed-methods approach…”*

The bibliometric and peer evaluator ratings data presented and its analysis are quantitative in nature. However, the authors refer to the ***process*** of rating by the peer evaluators as ‘subjective’, qualifying it a qualitative method.

### 4.1.10. Transition services for incarcerated youth: A mixed methods evaluation study. Abrams et. al (2008)

*Context and programme description*

The Transitional Living Programme (TLP) was a six-week intensive programme implemented by a public correctional institution for felony level juvenile offenders. The location of this institution is not disclosed. It was introduced in 2002 as part of the institution's comprehensive Intensive Aftercare Programme (IAP). The IAP was a 9 to 12 months correctional and rehabilitation programme where offenders were required to attend daily process groups and participate in a number of programs like cognitive-behavioral techniques, substance abuse treatment, vocational rehabilitation and other specialized therapeutic programs. Youth were housed in cottages and primarily sorted by age, gender, and level of risk, with the exception of one cottage designated specifically for male sex offenders. The TLP component focused on equipping participants with independent living skills. Participants spent the night in the TLP cottage and were released into the community during the day to attend work or school and gradually spend increased time over the weekends in their home settings. Case managers worked very closely with the youth to build their daily schedules and their plans for release. Only males were able to participate in the TLP intervention due to restrictions on mixed-gender housing.

*Description of the study*

The goal of the study is to describe and evaluate preliminary outcomes from the IAP-modeled programme. This goal is premised on the scarcity of information on how specialized services could facilitate successful transitions for youth reentering their communities from correctional placements. The specific aims are formulated as follows: -

1. To examine recidivism outcomes for youth participants in a transitional living programme at one-year post-release;
2. To explore child protective services involvement as a risk factor for recidivism at one-year post-release; and
3. To compare youth and staff perspectives on the strengths and limitations of the transitional living programme in preparing youth for community reentry.

The authors initially give a generic basis for using a mixed methods approach. They state,

> "*Mixed methods will be used to illustrate the benefits and limitations of this model in preparing youth for community re-entry*"

They later explain the role of the qualitative methods,

> *"…we used the qualitative component to understand and interpret the quantitative findings in more depth and with additional context".*

The methods used include a quantitative analysis of recidivism outcomes for graduates of the TLP cottage in 2003 and qualitative interviews with TLP youth participants and program staff from 2004–2005. The authors refer to the sequencing of the methods as a 'simultaneous implementation' of the qualitative and quantitative components though the qualitative interviews are conducted a year later. This could imply an understanding of independence of methods where the implementation of one is not contingent on findings from another.

The quantitative component addresses the first two aims and is a quasi-experimental design with participants in the programme selected using a convenience selection approach *(i.e. male offenders during the first programme year who accepted to participate)*, and non-participants (refusals) acting as the comparison group. The total number of participants is forty six. The comparison group comprises fifteen non-participants. Secondary data is the primary source of information. Descriptive analysis, bivariate tests and regression models are used to describe and compare the two groups with regard to demographics, outcome measures and possible correlations between variables.

The qualitative component addresses the third aim of the study. The results indicate that it is used to illustrate programme design and implementation issues. The selection of participants for the field study includes ten youth TLP participants, interviewed repeatedly over a six-month period, and one-time interviews with five TLP staff. A convenience selection approach is used for both categories. Twelve youth and five staff volunteered to participate in this component. Semi-structured interviews are conducted for the data collection. Questions are geared to gather staff perspectives on the important components of transition, the benefits and limitations of the TLP, and their views on the challenges of youths' post-release environments. Interviews with staff are completed in the correctional facility in a confidential meeting space.

*Results of the study*

The quantitative study reveals that there is no statistically significant difference between TLP males and non-TLP males on various recidivism outcome measures. From the thematic

analysis of the qualitative data, the following emerge: - Positive opinions about the 'refusal' and 'practical' skills obtained; that change is individual and largely "mental"; and some gaps /weaknesses of the programme which include: - negative opinions about the TLP concept, allocated time to the programme, and lack of follow-up for the youth.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The authors argue that the qualitative component of the study provided the context to understand the benefits and limitations of such a programme. From the quantitative component of the study, the authors infer,

> *"…this study found that participation in a six-week transitional living programme did not make a significant difference in recidivism outcomes at one-year post-release, and in this case, the TLP participants were slightly more likely to be reconvicted of offenses than non-programme participants".*

With reference to results from the qualitative component of the study, the authors comment,

> *"Despite unchanged recidivism rates, both youth programme participants and staff found the TLP to be highly beneficial in several ways. Youth reported gaining specific practical skills that support independence, especially with goals related to vocation and education. Youth and staff also believed that the opportunity to develop positive relationships with adults and to reinforce cognitive-based refusal skills in a supportive setting helped youth to better navigate the challenges of re-entry"*

The authors also note and explore the discrepancy between the quantitative and qualitative findings, stating,

> *"So what might explain some of the gaps between perceptions of the TLP benefits (by youth and staff) and its measured effects?"*

They conclude,

> *"...the main issue in this particular case was found in the lack of continued ties between the transition program and aftercare…"*

It is apparent that the authors use the qualitative findings and inferences thereof to illustrate the gap between the postulated proximal and distal programme outcomes. They use the discrepancy in the results about the two different program aspects studied by the respective

methods *(i.e. outcomes versus design/implementation)* to provide possible explanations for the absence of longer-term effects. They note that the qualitative component availed a context through which the outcome results were interpreted. Because the performance of the different programme aspects is logically linked, the design and implementation issues provide a validating function for the outcomes, which the authors could be referring to as the context for interpretation.

The authors also comment on the limitations of the quantitative component, particularly citing its non-experimental nature. Despite the many design limitations that they cite *(including sample size, duration of measurement, measurement construct explored, source of data and variables used),* they do not explore the possibility that the null effects revealed by this study component could have resulted from the inadequacy in the quantitative design. While the qualitative study indicates some benefits with regard to skills acquired, the negative perceptions about the programme concept and challenges experienced are more prominent and congruent with the inferences from the quantitative component. This seems to increase the trustworthiness of the inferences from the quantitative component despite the cited weakness in the design. It emerges from this discussion that the qualitative component plays a role of back-stopping the design weaknesses of the quantitative component. It is noteworthy that this role is realized through the focusing of the different methods on different and logically linked programme aspects.

### 4.1.11. A mixed methods evaluation of televised health promotion advertisements targeted at older adults. Berry et. al (2009)

*Context and programme description*

This programme was a public health television campaign targeted at adults in the state of Alberta in Canada aged 55–70 years. The goal of the campaign was to increase the number of Albertans who would be physically active and eat at least five to ten servings of fruits and vegetables per day. The campaign was launched in October 2007 and ran for eight weeks. It included two advertisements, one that encouraged consumption of fruits and vegetables and one focused on increasing physical activity within the target population. A central feature of the advertisements was a comical ‘‘grim reaper’’ character. The messages centered on small changes and a balanced life and a feature of the advertisements was to encourage viewers to

go to the *Healthy U* website for more healthy eating and physical activity ‘‘tips’’. All the advertisements were run on a local affiliate of the national, privately owned, television company.

*Description of the study*

This is an impact evaluation study that focuses on the design and effects of the programme. Three research questions guide the study, each with a corresponding data collection approach. The inquiry questions are based on a “Hierarchy of Evaluation framework” developed out of previous research in advertising and health promotion literature. The questions and their corresponding methods are formulated as follows: -

1. What is the campaign awareness, source of the advertisement, beliefs regarding either physical activity or healthy eating, intentions to visit a promoted website and intentions to perform the healthy behaviors *(telephone survey)*?
2. What physical activity and health eating advertisements are recalled by participants in general and therefore may represent competition for health promotion programmes *(answers to the unprompted recall survey question)*?
3. What were the positive and negative aspects of the advertisements, the perceived credibility of the source of the advertisements, and the usefulness of promoting a website *(focus groups)*?

It is apparent that the study bases the rationale for the mixed methods research approach on its ability to address the limitations of the typical methods used for such studies *(surveys and focus groups)*. The authors argue that by using both types of data collection, a richer sense of campaign impacts could be obtained. They define the role of qualitative component as providing depth. Specifically, the qualitative component is to help in the interpretation of the outcomes from the quantitative component. The focus groups are to be used or examining the advertisements in greater depth and to triangulate these results with the survey results to get a better understanding of the reasons the advertisements were or were not successful.

The study comprises two independent inquiries which are looking at different aspects of the programme. The quantitative component (telephone survey) focuses on effects (outcomes) while the qualitative (focus groups) focuses on an aspect of the programme efficacy (i.e. design of the advert). The focus groups are conducted after the survey, though the questions they explore are independent of the survey results.

The survey starts immediately following the end of the campaign and provides a cross-sectional examination of uptake of the senior specific campaign and lasted for a period of two months. The questions focus on: - recall of the advert (unprompted and prompted); Physical activity beliefs; intentions; Physical activity behavior; and demographics. Random sampling is used to select one thousand six hundred Albertans who participated in the survey with an oversampling of individuals older than 55 as they were the target audience for the advertisements. The analysis includes: - simple descriptive statistics, two logistic regressions and two analyses of variance (ANOVA). Codes are generated for the responses to the open-ended survey questions and frequencies of broader code categories for the physical activity and healthy eating generated.

A convenience selection approach is used to identify 29 community dwelling participants (aged 55 to over 80 years) for the focus groups. The focus groups take place a few months after the end of the campaign and "allow for a deeper exploration of perceptions of the campaign advertisements and considerations of advertising in general". The questions focus on perceptions of participants on the effectiveness of the advert in attracting their attention and causing them to act accordingly. Four focus group interviews are conducted in separate locations in the northern region of Alberta, Canada. A thematic analysis of the interview transcripts is done

*Results of the study*

For the quantitative study, the following results emerge - Unprompted recall was very poor; Prompted recall was somewhat better than unprompted recall; from the logistic regression analysis, significant results were obtained for only the age and level of education.

Three themes emerge from the qualitative analysis and include - a few positive but largely negative comments regarding the ''grim reaper" character of the advert; the government not being considered a credible source; and mixed reports about the use of the website and access to the Internet.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The authors draw a number of inferences from the two study components. In summary, the findings/inference made from the quantitative component is that "the exposure was inadequate and the campaign goal was not realized". The qualitative inferences indicate that the design of

the advert was lacking. The authors use the results from the qualitative study to explain the low impact revealed from the quantitative study. They state,

> *"Results of our qualitative data indicate that because of the distracting and to some, distasteful, nature of the ''grim reaper'' there would be little positive influence … as a result of watching these advertisements".*

The qualitative component which seeks to examine a specific component of the programme design in greater depth avail useful information for an explanation of the outcomes. It is evident from this study that a *'richer sense'* of campaign impact implies an interpretation or explanation of the impact findings/inferences.

The authors refer to a 'triangulation' of the qualitative and the qualitative results in a broader sense when compared to the stricter triangulation concept discussed in mixed methods research. In this study, the qualitative and quantitative components focus on different but related programme aspects, achieving a validating and explanatory function between them. It is apparent that the logical connection between the programme aspects *(perceptions about programme design, and outcomes)* compared facilitates such 'triangulation'.

### 4.1.12. Addressing the Challenges Faced by Early Adolescents: A Mixed-Method Evaluation of the Benefits of Peer Support. Ellis and Marsh (2009)

*Context and programme description*

The intervention was a peer support programme designed to assist students through the instability of adolescence and the transition to high school. The programme was conducted at three high schools located in New South Wales, Australia. The article does not mention the timing of the programme. The programme's overarching aim was to foster the physical, social, and mental well-being of young people. It aimed to do this by developing crucial values, skills, and attitudes that would not only assist students through the instability of adolescence and the transition to high school but also throughout their adult lives. More specifically, the programme aimed to assist students by enhancing the following key areas: - school competence, school citizenship, sense of self and possibility, connectedness and resourcefulness. The programme was designed to train senior year 10/11 high school students to work regularly with small groups of seventh grade students. Each group consisted of eight to ten Grade 7 students and two Grade 10/11 leaders. The leaders were responsible for directing the Year 7 students through

the programme content and activities. The programme consisted of twelve sessions, each of which was designed to run for 45 minutes and took place once per week.

*Description of the study*

The evaluation study has two aims - first, to test the impact of the programme on espoused programme outcomes and other aspects of students' psychological wellbeing and adjustment to the high school context that may have been affected by participating in the intervention; and second, to identify students' personal perspectives of the benefits of the programme. The rationale for a mixed methods approach is based on the ability of mixed methods to provide a comprehensive assessment of the programme. In this regard, the authors needed to address the two categories of study aims including - assessing impacts and understanding participant perceptions about the programme. They write,

> *"…recent developments which demonstrate that mixed methods studies can help elucidate various aspects of the phenomenon under investigation, providing a more holistic understanding of it, and resulting in better-informed recommendations".*

The authors articulate the following purposes for the respective qualitative and quantitative components. The purpose of the quantitative component is to test the effectiveness of the programme on espoused programme outcomes while the purpose of the qualitative study component was to enrich key quantitative findings, "giving voice and life to the results". The second aim of the qualitative component is to provide insight into issues that may not have been identified by the quantitative methods.

The quantitative component uses a longitudinal quasi-experimental design with a control group and baseline data against which to compare the intervention effects. The quantitative sample comprised all Grade 7 students (n=930) spanning a two-year period from three participating high schools. In the first year of the study, all Grade 7 students (n=478) from the three participating schools are assigned to the within-school (baseline) control group. In the second year of the study, all new Grade 7 students from the same three schools participate in the peer support programme and serve as the experimental group (n=452). The study has two qualitative components: - Open-ended questions as part of the questionnaire and focus group discussions exploring another aspect of the programme. For the focus groups, a random selection of 119 Grade 7 students and 44 peer support leaders are invited to participate.

*Results of the study*

The following results emerge from the quantitative analysis: - the following outcomes measures are statistically significantly higher in favor of the experimental group: - School Self-Concept, School Citizenship, Connectedness, Resourcefulness. The differences for the measure of Sense of Self and Possibility are not significant. From the quantitative analysis carried out on the open-ended questionnaire items, the majority of responses indicate a positive experience with the programme. The following themes emerge from the thematic analysis of the focus group discussions:- *Student Connectedness:* the programme helped strengthen student connections; *Problem-Solving Ability:* the programme enhanced Grade 7 students' problem solving and decision-making skills; *Sense of Self and Possibility:* the programme increased self-understanding and self-confidence; *School Citizenship:* the programme improved students' perceptions of bullying; *Adjustment to High School:* the programme helped students settle in and learn about high school. Results of both components indicate a focus on only the outcome/impact programme aspects.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The authors do not synthesize the qualitative and quantitative findings. They present the quantitative inferences and the qualitative results separately, with the quantitative inferences pre-dominating the discussion.

The initial intention of enriching key quantitative findings with qualitative findings is not very clear as there is no discussion linking the inferences from either method. Some qualitative findings *(i.e. 'Connectedness' and 'Sense of self and possibility')* that contradict with those from the quantitative component are not discussed. However, the purpose of providing insight into issues that may not have been identified by the quantitative method is realized in the various outcome themes that are not covered by the quantitative measures. The qualitative component whose inquiry focuses on programme strengths and weaknesses leads to three categories of results: - overlapping findings; contradictory findings; and 'new' findings. While the study takes up and discusses the 'new' findings, it is silent on and does not discuss the relationships between the qualitative and quantitative findings that contradict or overlap.

### 4.1.13. Using mixed methods to evaluate the Pediatric Lead Assessment Network Education Training program (PLANET). Polivka, et. al (2009)

*Context and programme description*

The Pediatric Lead Assessment Network Education Training programme (PLANET) was a peer-to-peer educational programme on lead poisoning in children developed in 2001 by the Ohio Childhood Lead Poisoning Prevention Program (OCLPPP) within the Ohio Department of Health (ODH). The target audience for the PLANET programmes was health care practitioners (physicians, NPs, nurses, dietecians, and medical assistants). The goals of the PLANET programme were to (a) raise awareness about pediatric lead testing among health care providers, (b) provoke action by distributing information about proper lead testing techniques, and (c) increase blood lead testing rates of at –risk children in Ohio. The three objectives of the PLANET programme were that attendees would be able to identify (a) the Ohio lead testing guidelines, (b) that lead poisoning will affect cognitive, social and physical development of children, and (c) the essential components used in educating parents and caregivers about childhood lead poisoning prevention. PLANET trainings were conducted and coordinated statewide through the OCLPPP's three Regional Resource Centers (RRC). Each RRC served as the main contact for all lead poisoning prevention in the region and provided case management support, culturally sensitive educational information, community outreach, referrals and presentation of PLANET programmes.

*Description of the study*

The aims of the evaluation study are formulated as follows: - to (a) identify how the PLANET programme was perceived by attendees and non-attendees and (b) determine the impact of the program in terms of aggregate changes in lead poisoning prevention attitudes and knowledge immediately after PLANET training and changes in physician behaviors specific to lead testing rates post-PLANET training.

The rationale for the mixed methods approach in this study is understood to emanate from the intention to use the qualitative methods to further explore the perceptions about the program. In describing the evaluation methods, the authors mention that these include existing data (from PLANET sign-in sheets, attendee-written evaluations, and pre- and post-tests) and conducting three focus groups (FGs) and twelve phone interviews (IVs) with PLANET attendees and non-attendees. A quantitative analysis of the written PLANET evaluations and qualitative analysis of

the FG and IV data are carried out. The quantitative component includes descriptive analyses of demographic data from the sign-in sheets and evaluation forms. The evaluators also review and categorize the written comments on the evaluation forms. A one sample t-test is used with the post-test as the study group and the pre-test as the population. In discussing the qualitative component, the study gives insight into the focus of the qualitative inquiry. The authors state,

> *"Both FG and IV participants described how they heard about PLANET, the setting in which the PLANET programme was administered, length of the programme, the PLANET curriculum, the use of the PLANET resource manual, impact of PLANET participation on lead testing practices, use of the blood testing guidelines, and methods to improve blood lead testing".*

There is an implication from the rationale of ***further*** exploration of perceptions using the qualitative methods that a sequential process is to be followed, where the FG and IV inquiry was to be informed by the results from the quantitative component. It is not clear whether the studies are sequentially or simultaneously conducted as the study does not mention when the quantitative component is conducted. However, an analysis of the quantitative results and the inquiry focus of the qualitative component give some leads. The quantitative findings focuses on participant demographics and opinions about the training and attitudes about application of the lessons learnt. From the above quote, it is apparent that the qualitative inquiry is not dependent on the quantitative findings but seems more of a separate inquiry into the training experience and its impact on them. It is therefore possible that the authors understand the rationale of "further exploration" as implying getting an alternative perspective or studying other aspects of the training programme using the qualitative methods.

*Results from the study*

The study presents descriptive statistics of - attendee demographics and attendee perceptions, which indicate overwhelmingly positive opinions of the training in particular and the programme as a whole. The study also presents a quantitative analysis of the open-ended questionnaire items (comments about the training), with most comments (90%) being positive. The FGs and IVs result in general recommendations concerning program delivery mechanisms and resources. The study alludes to some NULL effects with regard to change in behavior, noting,

> *"PLANET attendees reported that they did not necessarily change lead testing and lead education practices after the PLANET program although most acknowledged an increased awareness concerning lead poisoning".*

The authors also mention recommendations of attendees regarding the issue of educating of parents about lead poisoning and reasons by non-attendees for not participating. With regard to the programme impact, the authors state,

> *"…attitudes concerning the importance of lead poisoning and lead poisoning education as well as lead poisoning knowledge of attendees significantly increased after attending the PLANET program*".

They illustrate this with descriptive and inferential statistics indicating a statistically significant higher increase in blood lead testing among PLANET attendees in comparison to non-attendees.

*Discussion of the results/inferences from the qualitative and quantitative study component*

The findings and inferences are discussed and presented separately in the respective sections with no synthesis involved. It is only in the final discussion where the authors state,

> *"This independent evaluation of PLANET using mixed evaluation methods indicates that the program is positively received and is effective in improving blood lead testing rates"*

There is a contradiction between the qualitative and quantitative results with regard to PLANET attendee practices. The quantitative results indicate a significant increase in blood lead testing among PLANET attendees while the qualitative results show that attendees did not necessarily increase their practices after the PLANET training. The quantitative results indicate an increase in blood lead testing among non-attendees too. These three results point to a possible non-programme determinant of this outcome change. However, the authors are silent on the negative results from the qualitative component and only mention the quantitative findings in the final assessment.

### 4.1.14. A Mixed Methods Evaluation of the Effect of the Protect and Respect Intervention on the Condom Use and Disclosure Practices of Women Living with HIV/AIDS. Teti et. al (2009)

*Context and programme description*

The 'Protect and Respect' was an intervention designed to help women living with HIV/AIDS (WLH/A) decrease their sexual risk practices. Participants for 'Protect and Respect' were recruited from the Partnership Comprehensive Care Practice (PCCP) in Philadelphia, PA from April, 2004 through July, 2006 with women who were waiting for their regularly scheduled medical visits. The major activities included: - HCPs (nurses or physicians) attending an initial 4-h training and quarterly booster training sessions to learn the intervention's messages and prevention counseling skills; HCPs working with participants to assess their stage of readiness regarding condom use and/or disclosure, deliver prevention messages or teach skills, and create a follow-up plan; a health educator delivering the group-level intervention (GLI), which included five consecutive, weekly, 1.5 h sessions; and the study's Peer Educators conducting weekly 1-h support groups throughout the length of the project. Every woman who completed the GLI was eligible to attend the peer groups, which were designed to help women discuss the skills that they learned in the GLI in the context of their lives and challenges over the long term.

*Description of the study*

The study is an outcome evaluation of the intervention. The rationale for a mixed methods approach is premised on the need to make most of the strengths of either method. The authors state the inquiry components for which either method is used.

> *"Using both quantitative and qualitative methods allowed us to capitalize on the advantages of each method to obtain a more in-depth understanding of the effects of the intervention, as well as women's experiences within it".*

The study largely focuses on assessing the programme outcomes, with the qualitative component exploring aspects of the programme design/efficacy and implementation. The quantitative component is a quasi-experimental design. Participants are recruited purposefully *(women had to be at least eighteen years old, HIV positive for at least six months, and English-speaking)* over a two-year period. The recruitment also follows a convenience selection approach, where some potential participants are excluded because of being 'unapproachable', uninterested or not willing to participate. The recruits are randomly assigned to either a

comparison (N=81) or intervention group (N=77). Participants completed 'risk' surveys at baseline, 6-, 12-, and 18-months. This quantitative component is used to measure the two outcome constructs. The qualitative component is a field study and used to explore another aspect of the programme *(i.e. seeking to gain in-depth understanding of the experiences of women who participated in a component of the programme)*. It seeks responses to three questions: - Lessons learned; how the groups influenced positive outcomes; and the challenges in applying the skills acquired. A selection of participants (i.e. 18 women) from the quantitative study is recruited for the qualitative study. The interview questions focus on: - lessons learnt, help from the groups and application of skills acquired. The two studies are independent of each other, with the qualitative study following the quantitative.

*Results of the study*

The quantitative analysis reveal that: - there were no statistically significant differences between the treatment and control groups at any time point for the measure of disclosure. However, a higher increasing trend in disclosure for the intervention group is noted. For condom use, the statistical inferences are significant for two measures *(6 and 18 months)* in favor of the control group. From the thematic analysis of the qualitative data, five issues relating to opinions about the design/efficacy of the programme emerge. They include positive opinions about - the importance and relevance of the two components of the intervention design; the effectiveness of the peer support groups; support of the group facilitators and group dynamics; lessons and problem-solving skills. On the downside, participants in the qualitative study report two specific negative aspects despite these positive opinions: - (i) the difficulty in consistently integrating condom use into women's sexual activities and, (ii) that learning how to disclose did not always translate into actual disclosure.

*Discussion of the results/inferences from the qualitative and quantitative study component*

With reference to the contradictions that emerge between the qualitative and quantitative findings, the authors write,

> *"…the qualitative findings failed to explain fully why the intervention affected women's condoms use, but not disclosure, although it is also possible that the intervention's messages may not have been able to counter the risks and barriers…"*

From this statement, it is apparent that the authors use the inferences/findings from the qualitative study to try and explain the quantitative findings. For one outcome measure

(disclosure), they prefer to use the trend statistics as opposed to the significance tests possibly because the former are congruent with while the latter contradict with the qualitative findings. The authors discuss a number of limitations in the quantitative measures and its design as possible reasons for the lack of statistical significance. Additionally, there is a suggestion by this statement that the qualitative findings explain the increases in disclosure among the intervention group when compared to the control group. However, this seems not to agree with the qualitative finding that "learning how to disclose did not always translate into actual disclosure". This contradiction therefore seems to affect the ability to avail a complete explanation as it is not clear how the increases in disclosure came about. For the other outcome measure (condom use), the authors note that the qualitative findings failed to explain the paradox *(the intervention appeared to have adverse effects on the participants)*. With reference to the questions defined and results obtained from the qualitative study, it is apparent that the qualitative study may not necessarily have been designed to explain this quantitative finding. The fact that the qualitative study does not make reference to the results from the quantitative results in its design would at best provide a speculative explanation but would typically preclude the possibility of an effective and logical explanation.

## 4.2. Discussion

In this section, a number of themes resulting from the analysis of the various studies presented in the previous section are discussed. The aim is to synthesize the findings across the studies within the framework of the study components defined in the analytical framework. The first part of the discussion presents categorizations within the respective study components while the second part traces common themes between or across the different components.

***Rationales for the use of a mixed methods approach***

A number of different but not mutually exclusive rationales are given by or inferred from authors' statements for the use of a mixed methods approach. For some studies, a number of these are included in the discussions. However, it is noteworthy that one category of studies does not mention any reason but just describes/qualifies the research approach used as a mixed methods design. Some of these authors refer to a combination of the approaches,

> *"The combination of these two approaches is sometimes referred to as mixed methods"*

while others refer to the use of multiple methods in the study,

*"…a mixed-method process evaluation using multiple data sources with triangulation of quantitative and qualitative data was conducted".*

*"The approach taken in this pilot evaluation is aptly described as mixed-methods…because multiple qualitative and quantitative measures and analyses were incorporated into the design".*

For some studies in which the primary research approach has an emphasis on quantitative data, the rationale for a mixed methods approach is based on the need to obtain an alternative perspective to the programme using the qualitative component.

*"Interviews were conducted with young people….in response to requests from programme teams for more qualitative components in the evaluation plan".*

*"The goal of the content analysis…was to understand…what was the perceived impact of the intervention on the facility and staff from the perspectives of the managerial informants".*

A closely related group premises the rationale on the ability of a mixed methods approach to give a more comprehensive picture of the evaluand, qualifying it as the ability to capture sufficient information and increasing the validity of the inferences made.

*"Using mixed methods allows for a more comprehensive approach to evaluation by enhancing both the ability to capture adequate information about the phenomena under study and the validity of data interpretation".*

Another set of rationales is based on the need to collect multiple data. With reference to the extensive requirement of the evaluation agreement and the need to include qualitative and quantitative data, these authors state,

*"The use of multiple methods was necessary because of the many types of information required…"*

For another category of studies, the rationale is the use of a qualitative component to complement a primary quantitative evaluation method/approach. The complementary role includes aspects like understanding, interpreting and enriching the quantitative findings. For some of these studies, the complementary role is stated in a generic way,

*"…we used the qualitative component to understand and interpret the quantitative findings in more depth and with additional context".*

*"…qualitative methods were used to …strengthen our understanding of the quantitative results"*

*"…the purpose of the qualitative study component was to enrich key quantitative findings".*

while others are specific on the focus of the complementary role,

*"The focus groups were to be used or examining the advertisements in greater depth…to get a better understanding of the reasons the advertisements were or were not successful*

Yet another category of rationales emphasizes the need to address the inherent weaknesses within and make most of the strengths of the various qualitative and quantitative methods.

*"Using both quantitative and qualitative methods allowed us to capitalize on the advantages of each method to obtain a more in-depth understanding of the effects of the intervention, as well as women's experiences within it".*

Some of these studies refer to the weaknesses in either method,

*Mixed methods research can overcome the limitations presented by the two most popular methods used in such research, surveys and focus groups*

while others seem to emphasize the weakness of the quantitative method only.

*"The second aim of the qualitative component was to provide insight into issues that may not have been identified by the quantitative methods".*

*"Quantitative methods… may not provide the information that program planners and decision makers need …"*

***The understanding of a 'method'***

There are a few variations in the way the studies conceive what a method is. For one category of studies, methods refer to data collection (e.g. focus groups, interviews, questionnaires) and

the related analysis (thematic, descriptive, and inferential). For another category, it is conceived at the level of methodologies (e.g. surveys, field studies, bibliometric analysis, concept mapping) while for another, it is a mix of methodologies and data collection (e.g. survey and interviews). Another group has a seemingly broader conception (outside the research methods norm) of what qualifies as a qualitative method. These include the 'informal' processes involved in research like the preliminary "discussions" that inform the evaluation plan,

> *"The first round of interviews with key service providers served to initiate contact with many service providers and raised awareness of the programme and its aims. This was also the case in the second round of interviews".*

> *"In this study we started with the qualitative question, "What should be evaluated?" which was followed by a quantitative assessment of the responses to the questionnaires.*

and the 'subjective' process of collecting quantitative data,

> *"For instance, we gathered information about publication quality that included quantitative bibliometric data on citation rates and journal quality, and subjective ratings from center researchers and multiple peer evaluators.*

***The nature of results generated within a single study***

The nature of results generated within a single study fall under three groups. One has both 'qualitative' and 'quantitative' results. This includes studies which contain both qualitative and quantitative data collection instruments and those that have single or multiple 'qualitative' and 'quantitative' research methodologies. The second group of studies has only 'qualitative' results, and involves the use of 'qualitative' methodologies, methods and data in the core component of the research (i.e. case studies and qualitative data collection) only, with the quantitative component used within the sampling process to select cases for study. The third group has quantitative results only and includes: - the collection of quantitative and qualitative data using a 'quantitative' instrument, but having only quantitative analysis of both data sets, and multiple quantitative methodologies collecting only quantitative data, with the qualitative aspect of the study being the 'subjective process of determining the quantitative ratings.

***The assessment focus of the qualitative and quantitative methods***

The inquiry focus of the respective qualitative and quantitative methods is also explored. One category of studies uses the quantitative method as the primary assessment method for the

effectiveness of the program. Under this, a sub-category uses the qualitative method for a complimentary role to study various programme aspects which include: - design/efficacy, implementation or a combination of these. Another sub-category includes qualitative methods which focus on assessing outcomes and is related to the mixed methods rationales of: - obtaining a more comprehensive picture, a qualitative perspective or addressing the limitations of the quantitative method. Another category of studies uses both the quantitative and qualitative methods to assess programme design and implementation. Studies with this focus are congruent with the mixed methods rationale of collecting multiple qualitative and quantitative data. Largely, the qualitative and quantitative methods are concurrently conducted while for a few, the qualitative follows the quantitative. It is common to have the quantitative method as the primary evaluation approach focused at producing predominantly descriptive results on outcome/impact measures. On the other hand, the qualitative component is considered to play a complimentary role which includes issues like: - providing an in-depth understanding through the programme design, processes and context, and helping to clarify and interpret the quantitative findings.

***The synthesis of findings/inferences from the qualitative and quantitative results***

The studies vary in the way they discuss the qualitative and quantitative results and findings towards richer evaluation inferences. One category of studies does not include a synthesis of the qualitative and quantitative results/findings. In such studies, the respective results are presented on their own and never combined or related. In a number of studies falling under this category, the authors are silent about the discrepancies and contradictions between the qualitative and quantitative results. In these cases, the authors typically prefer the positive results (this includes both qualitative and quantitative) which they emphasize in their conclusions.

A group of studies do slightly more *(though not necessarily qualifying as a 'synthesis')* than that discussed in the previous category by presenting the qualitative and quantitative results that fit under a specific theme together. These studies present one method results first and then follow it up with the alternative method results without any synthesis. The quote below from one of such studies refers.

> *"In this evaluation, the quantitative method used…yielded information about infant mortality rates … and prevalence of inadequate weight gain…. In contrast, the use of an*

*"anthropological approach" yielded information about health decision–making processes …"*

Another set of studies explains the discrepancy/contradiction between the qualitative and quantitative findings as resulting from design limitations particularly from the quantitative method.

*"The reasons the quantitative findings do not detect what appears to be positive change…but may be due to a number of factors, including a sample size that was too small to assess anything but very large changes; a follow-up timeframe that was not long enough to evaluate change appropriately; or inappropriate outcomes or survey instruments".*

A common feature among such studies is that the qualitative results are positive while the quantitative results are negative. It is apparent that the authors prefer the qualitative results or are sure of the validity of the qualitative design and use them to point out design limitations of the quantitative method. A variation in this approach is a case where the authors opt to use the positive descriptive statistics as opposed to the negative inferential statistics that are not congruent with the qualitative findings.

Another category of studies synthesizes the qualitative and quantitative findings/inferences to generate richer inferences. For this category, the qualitative method plays a complementary role and the authors discuss the congruence of the respective results, the completeness of the findings, or how the qualitative findings explain the quantitative findings.

***The relationship between the rationales, methods focus and the synthesis of findings***

Two sets of relationships between the rationales given by the authors and the way they synthesize the qualitative and quantitative findings are noted. The first focusing on rationales includes: - no rationale at all, the collection of multiple data (qualitative and/or quantitative), and those where authors intend to 'address the weaknesses' of one or either method. This category also includes studies in which an intended complementary role of the qualitative method is not actually implemented. For all these cases, there is largely no synthesis of results and in most cases, the authors are either silent about or explain away results that contradict their preferences in making the final conclusions. A common feature among the studies with contradicting results in this category is that the alternative methods focus on assessing the

same programme aspect (particularly effects/impact) and are independent of each other (conducted concurrently). In discussing the final results, the authors are silent about the non-preferred findings when they fail to reconcile the contradictions or explain them away when they do. A related common feature is that the complimentary role of the qualitative component is not clearly defined. It either lacks any guiding research questions/aim and for those where it is mentioned, it is too broad/generic. This could possibly explain why some studies with an intention of using the qualitative methods to study a programme aspect different from that of the quantitative method end up with results focused on the same programme issues. In many cases, it leads to results that are already captured under the quantitative component. Some of these studies have congruent qualitative and quantitative results/findings which would potentially improve the validity of the respective findings but are not discussed possibly because it is not intended. This includes some seemingly possible explanations for non-effectiveness which are apparent from the results about the programme design and implementation. Some studies in this category try to explain the quantitative findings using qualitative data. However, the explanations are not consistent with the data/results, being based on speculations.

The second category includes studies with rationales and methods use where the qualitative component plays a complimenting role to a primary quantitative method. In these studies, the qualitative and quantitative findings are synthesized to provide more insightful and richer evaluation inferences. A common feature among these studies is a planned approach to the role and use of the qualitative method. Though the methods have both a concurrent and sequential conduct of the studies, the question/aim of the qualitative method and how it is to compliment the quantitative method is very clear and specific. The two quotes are illustrative.

> *"The goal of the focus groups was to examine the advertisements in greater depth and to triangulate these results with the survey results to get a better understanding of the reasons the advertisements were or were not successful".*

> *"The second part of the study was a more qualitative study. Here, the objective was to further analyze the reasons for the different attitudes on the wards. The quantitative study exactly described these attitudes, and the qualitative study was now intended to further explain those quantitative results".*

It is worth mentioning that for studies in this category, the qualitative and quantitative components focus on different programme aspects (design/implementation versus outcomes

respectively). Within these studies, there are both congruent and contradicting findings. However, unlike for the previous category, the contradictions play a complimentary role by explaining the findings from the quantitative component. Similarly, the qualitative findings that are congruent with the quantitative findings improve the reliability or trustworthiness of the quantitative inferences.

***The relationship between the sequencing of the qualitative and quantitative methods and the synthesis of findings/inference***

The case studies include both sequential and concurrent implementations of the qualitative and quantitative methods. Within the sequential implementation, the qualitative study follows the quantitative one, its inquiry approach being driven by the findings/conclusions from the quantitative study. For the studies with a concurrent methods implementation, the qualitative and quantitative inquiries are independent of each other, though the timing has both simultaneous and follow-up options. Among the studies with a sequential implementation, the qualitative inquiry is driven by a specific inquiry objective/question and the results are used to avail explanations for the quantitative findings. The studies with a concurrent methods implementation have two variations with regard to the qualitative component. The first includes those with specific aims/questions on the complimentary role of the qualitative method. These studies seem to hypothesize the possible results/findings from the quantitative component which they use to inform the qualitative inquiry. It is noteworthy that such studies synthesize the findings from the two components. The second category includes studies in which the qualitative component is not introduced within a context of complimentarity but as another study method. In a number of these studies, the qualitative and quantitative results are not synthesized. Some few use the qualitative findings/inferences to explain the quantitative results. However, as acknowledged by some authors, the explanations are partial and largely speculative, i.e. they are not based on the qualitative data included in the study. The statements below illustrate: -

> *"The ability of the intervention to increase WLH/A's condom use at months 6 and 18 may have resulted from the intervention's multiple and reinforcing safer sex messages…"*

> *"By delivering consistent but varied messages this approach may have given the women sufficient knowledge…"*

*""The qualitative findings also indicated that patients engaged in the program by…which could have affected the outcomes of their experience".*

*"Part of the improvement in perceived quality of care…may have resulted from the coaching supervision component".*

Among this last category of studies, some authors have intentions that imply a sequential conduct of the methods *(with the qualitative component complimenting the quantitative in a follow-up design)* but implement the methods concurrently. Some statements made by these authors include,

*"The purpose of the qualitative study component was to enrich key quantitative findings, giving voice and life to the Results…, to ''put flesh on the bones of the quantitative results…"*

*"Perceptions about PLANET were further explored via FGs/IVs…"*

*"Qualitative methods were used as a follow-up to the conclusions from the ecological data analysis…"*

It is apparent that such intentions cannot be effectively realized with a concurrent implementation of the methods, and may explain why some are challenged when they try to synthesize the findings, particularly in using the qualitative findings to explain the quantitative results.

***Vagueness and discrepancies between authors' assertions and implementation***

A common trend noted among a number of studies is the discrepancy and vagueness between the assertions made by authors and actual practice reflected through the way the methods are implemented. This is particularly among studies where the qualitative methods are proposed to compliment the findings from the primary quantitative component. The discrepancy commonly relates to the reference to a sequential use of the methods, with the qualitative method used in a follow-up approach to the findings from the quantitative component. The typical statements made are reflected as intentions or conclusions and include,

*"…the purpose of the qualitative study component was to enrich key quantitative findings, "giving voice and life to the results"*

*"Qualitative methods were used as a follow-up to the conclusions from the ecological data analysis…and supplemented the quantitative research…"*

*"…qualitative data in the form of focus groups and interviews were collected midway through the follow-up period to further explore program impacts"*

There is a suggestion in these statements that the qualitative inquiry would be informed by findings from the quantitative component. However, among all such studies, the inquiry focus of the qualitative component is independent and focuses on relatively different programme aspects. Consequently, none of the studies endeavor to synthesize the findings to support the assertions made, with the qualitative findings making no reference at all to the quantitative findings and vice-versa. While such a discrepancy could be argued as being a misinterpretation among those studies in which the assertions are intentions, it is more evident among those where such assertions are made in the concluding remarks.

It is also noted that the timing of some of these studies is sequential, with the qualitative component following the quantitative. It is therefore possible that these authors understand notions of "follow-up" and "further exploration" more in the sense of timing as opposed to methods dependency.

The vagueness is evident within the statements about the methods uses and the achievements made in the concluding remarks which are not evident in the description of the methodology. Typical statements include,

*"…there were also elements of complementarity…where quantitative information was used to clarify and contextualize quantitative results".*

*"Using a mixed-methods evaluation minimized the weaknesses inherent in each method…A mixed-methods evaluation permits triangulation of data and is a promising strategy for identifying best practices"*

*"…the mixed-method approach to the evaluation enhanced the ability to interpret the quantitative findings and provided a broader understanding of patients' perspectives than a quantitative analysis alone".*

These assertions are neither evident in the way the methods are implemented nor in the discussions of the findings. Additionally, the authors do not endeavor to illustrate or elaborate on them. Similar to the previous category, none of such studies synthesize the findings from either method.

***Quantitative findings/inferences and the way studies generalize findings***

For a number of the case studies, the sampling/selection approach is a combination of convenience and purposive selection. For many of them, the numbers are small and therefore data collection targets the whole population. Some studies, through the use of descriptive statistics, demonstrate that the sample is representative of a larger population. It is noted that none of the studies take these inferences further to represent a wider population. They largely characterize the study sample and are more interested in picking lessons of how similar programs would be implemented elsewhere. Examples of such characterization include: - proposals for better program designs, description of how the program influenced change and how that happened, and limitations to effectiveness. It is apparent that the concern of transferability of findings is more relevant to these evaluations in preference to generalizing to a larger population as most study discussions emphasize program improvements and/or recommendations in case of program replication. A common trend noted in many of the studies is that most statistical inferential tests indicate non-significant effects while the qualitative themes indicate benefit on the same measure. In one case, the descriptive statistics indicate positive results while the inferential statistics give negative results. Apart from a few surveys with large numbers (>100), many of the other surveys have non-significant differences between experimental and comparison groups for many of the measures. Similarly, tests to prove that the comparison and control groups are similar are non-significant. It is possible that the small sample sizes erroneously affect the potential to detect any differences, right from the characteristics of the experimental and comparison groups.

In summary, two broad themes emerge across the various categorizations identified above – a weak and a strong conception/implementation of mixed methods evaluation. In the weak conception of mixed methods, a synthesis of the findings from qualitative and quantitative components is very uncommon. In many cases, the respective results are presented separately with neither making reference to the other. In such studies, even apparent congruencies or contradictions between the results from the qualitative and quantitative components are not discussed. In a few, where a synthesis is attempted, contradictions between results create a

challenge for the authors. In some instances, the authors are either silent about these or exclude the negative results, while for others, they give explanations for the preference of the qualitative results over the quantitative, citing design weaknesses particularly of the quantitative method.

Such contradictions seem to emanate from two sources: - a lack of specificity about the inquiry aim for either method (but typically the qualitative component), leading to a collection of qualitative and quantitative data on the same programme aspect; and the use of inferential statistics for the quantitative component – a measure that does not seem to effectively elicit the effectiveness of some programmes. The primary characterization of this weak conception of mixed methods evaluation seems to be the lack of an a priori intention of integrating the findings/inferences from the qualitative and quantitative methods used. The inquiry approach or the rationales given for a mixed methods approach do not call for a combination of the qualitative and quantitative methods but typically focus only at the data collection level. This includes categories that: - do not give any rationale for using both methods; those which justify the use of mixed methods on the basis of the need to collect multiple qualitative and quantitative data; studies with rationales aimed at addressing methods weaknesses but not necessarily the results thereof; and studies that seek 'comprehensiveness' or 'an alternative perspective' without being specific on how an integration of the results/findings is supposed to contribute towards this. It is noteworthy within this weak conception that authors wrongfully assert benefits of a mixed methods approach despite lack of any evidence of this in the way the research is conducted or in the presentation of the results.

There is an implication from most of these rationales that the collection of multiple qualitative and quantitative data within a single study qualifies it as a mixed methods approach. With regard to the methods used, this weak conception qualifies processes like - preliminary discussions with stakeholders that inform the evaluation plan and the 'subjective' process involved in research as qualitative methods. Others qualify the quantitative approaches (e.g. ranking) used in the sampling/selection process. Typically, these studies do not necessarily collect both qualitative and quantitative data but are either predominantly quantitative or qualitative in nature. Additionally, the methods are independent of each other and are conducted concurrently.

Within the strong conception of mixed methods, the findings/inferences from the qualitative and quantitative components of the study are synthesized, leading to richer evaluation inferences. A common feature among the mixed methods rationales given in these studies is the definition of a clear role of both the qualitative and quantitative methods (but particularly the qualitative method), and specificity about how the findings/inferences from the respective methods are expected to contribute towards richer evaluation inferences. It is clear in the statement(s) of these rationales that some form of integration of the findings from the qualitative and quantitative components is required. The rationales falling under this category include those in which the complimentary role of the qualitative component is clearly and specifically defined. They include aspects of both sequential and concurrent conduct of the included methods. Within this strong conception, what qualifies as a 'method' is either a research methodology or a data collection/analysis method. Each of these methods includes a complete research process of sampling, data collection, analysis and presentation of findings/inferences. Another common feature among these studies is their focus of the qualitative and quantitative methods on different programme aspects, which provides for both congruence and completeness/comprehensiveness of evaluation inferences.

# Chapter 5 – Conclusions and recommendations

This thesis is premised on the contention by various researchers that evaluation studies by their nature employ a multiplicity of methods and therefore stand to benefit from the use of mixed methods research approaches in their designs. The rationale for the study is based on the argument that while a lot had been written about the use of mixed methods in evaluation, there is a dearth of literature that focuses on how the use of a mixed methods evaluation approach facilitates the realization of richer conclusions or inferences about programme merit/worth. The research question set out to explore how authors who foreground a mixed methods evaluation approach in their studies have interpreted it and consequently used the qualitative and quantitative components in the making of evaluation inferences. An analytical framework is developed on the basis of a literature review on mixed methods research and the background literature on evaluation research. This framework guides the qualitative content and thematic analysis of fourteen case studies. The results of the analysis reveal two major categories of approaches relating to how the interpretation of the notion of mixed methods evaluation influences the way the qualitative and quantitative methods are used and how the inferences are made. In the sections below, the implications of these findings to evaluation research are discussed with reference to previous research. Consequently, some lessons and recommendations for the novel approach of mixed methods evaluation are made.

Proponents for mixed methods evaluation have justified it on the basis of two premises: - (i) providing valid and quality data/knowledge and better understanding, and (ii) allowing for multiple, diverse ways of knowing and valuing, hence supporting an all-inclusive/democratic perspective about the performance of a given social intervention. Greene, et. al. (2001) elaborate on these categories through two perspectives in their discussion of the meaning of "better understanding" achieved through the use of mixed methods. More details on the congruence of these two perspectives with the two major study findings are discussed in the following sections.

The first perspective blends with the former premise and views "better understanding" as: - *'enhanced validity and credibility of findings'* through triangulation to rule out the threat to validity; and '*more insightful understanding'* where non-convergent or conflicting results lead to new insights and hence further explorations about the phenomenon. It is congruent with the 'strong' conception of mixed methods evaluation defined in this study which is characterized by mixed methods purposes of: - 'complementarity' which seeks to measure overlapping but also

different facets of a phenomenon, yielding an enriched elaborated understanding of that phenomenon; and 'triangulation' which seeks congruence. The primary common feature among studies falling under the strong conception of mixed methods is the synthesis of findings/inferences from the qualitative and quantitative components of the study. It is apparent for this category that the planned integration or combination of the qualitative and quantitative components, guided by a clear rationale for the integration of the methods/results sets the stage for an effective synthesis of the findings/inferences. This approach to a mixed methods study conforms with the proposals of Tashakkori and Creswell (2007) and Onwuegbuzie and Leech (2006) who argue that a mixed methods research study needs to start with a mixed methods question (embedding both qualitative and quantitative questions) whereby the questions dictate an integration of different aspects of the alternative methods used. This perspective also resonates with the definitions and designs of mixed methods research that emphasize the combination/integration/mixing of the qualitative and quantitative methods, data or findings.

The second perspective blends with the premise of multiple methods use and views "better understanding" to mean: - 'greater comprehensiveness of findings' where the lenses of different methods are focused on different aspects of a phenomenon to provide a more complete and comprehensive account of a phenomenon; and 'Increased value consciousness and diversity'. This conception of mixed methods evaluation seems to equate *multiple methods research* to *mixed methods research*. A number of evaluation researchers have proffered this perspective including: - Greene (1997) who justify mixed methods in evaluation on the premise that evaluation practice is characteristically pluralistic, embracing diverse perspectives, methods, data and values, therefore requiring a multiplicity of methods; Chen (2006), Rallis and Rossman (2003) and Madison (2002) who premise the application of mixed methods on "…the pragmatic needs of answering evaluation questions...". Study designs falling under this category have the intent of seeking to extend the breadth and range of inquiry with a rationale of increasing the scope of inquiry through the selection of the methods most appropriate for multiple inquiry components. Particular attention is drawn to case studies with designs that follow a mixed methods purpose of 'expansion'. A number of these studies are consistent with the 'weak' conception of mixed methods evaluation whose primary characterization is a lack of any synthesis of the findings/inferences from the qualitative and quantitative components. The few that try are either challenged when contradicting results emerge or give incomplete or speculative inferences. The common denominator of studies in this category is the absence of an explicit intent or plan of integrating the methods/data/findings from either method. It has been

illustrated from this study that the multiple methods used and/or data collected typically lead to multiple inferences and not inferences resulting from a mix of the different methods. Definitions of mixed methods research that focus only at the use or inclusion of multiple methods or data conform to this perspective.

An interesting pattern is noted among studies that seek congruence between findings from the qualitative and quantitative components or have a triangulation purpose. The findings indicate that a focus of both methods on the same programme measures (e.g. outcome/impacts) creates many challenges including failure to reconcile contradictions between either findings/inferences to irrelevance/redundancy of results/inferences from either method. This is congruent with the findings by Morgan (1998) who argues that the goal of triangulation is a rare motivation for combining qualitative and quantitative methods, referring to the impasse that arises when results from different methods fail to converge and the effort that goes into the "fairly limited goal of producing convergent findings". On the other hand, a focus of the methods on different aspects of the programme (e.g. qualitative for processes/design and quantitative on outcomes) serves both a complimentary and corroborating function for either method/results. Though not having the ideal triangulation design, the qualitative inferences about one aspect of the programme (e.g. design, processes) add to the validity or trustworthiness of the quantitative inferences about programme effects. It emerges that a design with a corroboration purpose is more effectively and practically realized when the qualitative and quantitative components are used to study different but related programme aspects like efficacy, efficiency and effectiveness. Because the performance of the different programme aspects is logically related, the inferences of the various individual programme aspects have a mutually validating and enriching function which can be harnessed when either method is focused differently. This is in agreement with findings by Morgan (ibid) who notes that the purpose of complementarity has been the most common alternative to convergence, proffering an approach based on decisions of method priority and sequencing.

An insight from this discussion is that evaluation studies can leverage the potentials of a mixed methods approach within a context of a comprehensive assessment. Such a context requires the assessment of different but related programme aspects, with the respective methods used for an aspect they are most appropriate to measure. For example, the results and inferences from the qualitative methods used at the design and process evaluation stages play a complimentary and validating role for findings for both the quantitative and qualitative

components at the outcome /impact level and vice-versa. This approach to the use of the qualitative and quantitative methods can also be leveraged in addressing the causality concerns in evaluation. The growing prominence for explaining effects as opposed to just describing them can harness the causal explanatory potential of qualitative methods, logically linking the programme conception, inputs, implementation, and outputs to the outcomes or impact.

This thesis defines a richer inference as that which emerges from a reflection on or synthesis of the data, findings and/or inferences from both the qualitative and quantitative components of the study. While the inferences made by studies falling under the 'strong' conception conform to this definition, those under the 'weak' conception fall short of making richer inferences defined within this context. With reference to the findings of this study, it is argued that when a research approach does not intend to or have a planned approach of integrating the qualitative and quantitative methods used, it will most probably not be able to synthesize the findings/inferences effectively (if at all). It will therefore be unable to make any richer inferences beyond the multiple inferences from the different methods. This contention has implications for studies that justify the use of multiple methods on the premises of: - 'a more complete' and 'more comprehensive' account of a phenomenon, 'allowing for multiple diverse ways of knowing and valuing', 'supporting an all-inclusive/democratic perspective' and others availed through multiple methods research. It is argued that such justifications could be trading inference quality/richness achieved through an integration of methods for inference quantity. On the contrary, as has been illustrated from a number of the case studies, multiple 'independent' inferences can potentially lead to a poorer and less valid overall impression about an evaluand by weakening each other, especially when there are contradictions between the multiple findings from the qualitative and quantitative components.

Studies in which the findings from the qualitative and quantitative components contradict each other reveal another interesting insight with respect to the thinking of authors about the roles of the qualitative and quantitative traditions. The common and pre-dominant designs among these studies are a combination of quasi experiments or longitudinal studies with pre-post measures for assessing outcomes/impacts and various qualitative methods that play a complimentary role. The results indicate that the findings/inferences from the qualitative and quantitative components contradict, especially where the respective methods focus on the same programme aspect – particularly outcomes/impact. While such discrepancies are potential sources for insightful understanding through further inquiry, the authors fail to reconcile them. Subsequently,

some authors explain the discrepancy as resulting from design limitations of the quantitative component while others prefer inferences from one method over the other, largely ignoring the alternative method inferences. It is noteworthy that in a number of these cases, the inferences from qualitative components are preferred, even when it is the complimentary method to the primary quantitative evaluation approach. This is especially so when the findings from the quantitative component are negative while those of the quantitative are positive. It seems that positive findings overrule negative findings, irrespective of how they are generated! The positive qualitative findings are seen as shedding light on neutral/negative/positive quantitative findings, i.e. the qualitative results are used to explain gaps or discrepancies or NULL results on the quantitative side. Similarly, it seems like the study is kind of "saved" by the qualitative component whose positive results are invariably taken at face value (they are more truthful than the negative results from the quantitative component). It is therefore argued that studies of this nature reduce the enriching purpose of a mixed methods approach to one of using one method (particularly the qualitative) to 'rescue' the overall study. The foregoing discussion reveals a possibly limited and narrow purpose of using a mixed methods approach among some practices of evaluation.

The concerns of Giddings (2006) regarding the philosophical paradigm of mixed methods research as "positivism dressed in drag" also seem to resonate within this set of case studies. As noted in the foregoing paragraph, the pre-dominant designs are a combination of quasi experiments or longitudinal studies, with the qualitative methods playing a secondary complimentary role. Though not mentioned or acknowledged in any of the studies, it is possible that these evaluation researchers approach the inquiry with a 'quantitative' mind, underscoring the argument that quantitative approaches may still be preferred to and more acceptable than the qualitative alternatives as the primary assessment methodology.

The ability to make richer inferences through a synthesis of findings/inferences from the qualitative and quantitative study components is contingent on the "truthfulness" of the findings from the respective components. A meta-inference based on a synthesis of inaccurate inferences will mislead the final assessment, possibly being an even weaker inference than the individual inferences alone. Some evaluation researchers have suggested that some study designs do not effectively elicit programme effect, or erroneously lead to wrong inferences. Perry, et. al (2007), illustrate that a standardized quantitative model for assessing a school outcome construct - "Academic self-efficacy" fails to elicit programme value which they

demonstrate is very evident with the alternative qualitative measure. She cites a number of other studies which reveal the same findings. While this could be true for both qualitative and quantitative methods, the findings from this study suggest that it particularly relates to the standardized impact/effect assessment quantitative models which dominate the effect/impact assessment designs. A case in point is one of the studies in which the descriptive statistics from the quantitative component reveal positive results congruent with the findings from the qualitative component while the inferential statistics are contrary to this. The suitability of quantitative approaches in assessing programme value is elaborated upon from a valuing and a design dimension. From a valuing perspective, some evaluation researchers like Eisner (1994, cited by Alkin, 2004) argue that what matters cannot be measured quantitatively. Others like Stake. et.al (1997) and Stake and Schwandt (2006) propose to enhance the process of explicating an evaluand's value through experiential knowledge, which resonates more with a qualitative approach. From a design dimension, limitations of the popular quantitative designs particularly for programme evaluation research have been noted. With reference to the measure of 'significance testing', Neil (2008) argues that significance tests should only be used when generalizing from the sample to the population. He notes that typical evaluation studies do not have the need or interest to generalize findings. This is underscored by the findings on the inferences of the cases in this study which demonstrate that the studies seek to describe the specific programme impact and possibly its improvement as opposed to generalizing the findings. He therefore argues, "In these situations, descriptive statistics and effect sizes are all that is needed". Neil adds that significance tests conducted with low power can be misleading, leading to Type II errors (incorrectly failing to reject the null hypothesis). He notes that programme evaluation studies with less than approximately fifty participants tend to lack sufficient statistical power. He writes that significance tests conducted with high power can be misleading, citing studies involving large sample sizes (e.g., > ~ 400), where even trivial effects are likely to become significant. He surmises by recommending the use of effect sizes with confidence intervals as opposed to significance testing. In the same vein, Schmidt and Hunter (1995, cited by Lalithamma and Khosravi, no date) caution against this "all too common error" when reporting non-significant results. They contend that the inference by some authors that non-significance implies no effect is incorrect, arguing, "… a non-significant result only indicates that the data being tested do not provide adequate evidence to reject the null hypothesis, given a particular alpha level. The non-significant result does not demonstrate that the null hypothesis is true". Therefore researchers planning to follow a mixed methods evaluation approach need to

be cautious by verifying the suitability of some 'popular' quantitative designs in eliciting programme value.

In summary, this study has explicated two prominent perspectives of mixed methods evaluation prevailing among evaluation researchers and practitioners. The first has the intention of and places emphasis on the integration of the qualitative and quantitative components *(methodologies, methods, data, results and inferences)* with the primary objective of obtaining richer evaluation inferences, hence addressing the 'technical' needs of evaluation. In this conception, the use of the methods and the data/inferences thereof are synthesized to achieve this goal. It is noteworthy that this conception is not much different from the classic 'mixed methods' studies that characterized the first half of the twentieth century and the multiple methods and triangulation designs that followed. Both approach the inclusion of multiple methods and data in a single study with an intention of integrating the data or findings for richer and more valid inferences. A multiplicity of methods and the integration thereof is perceived as good and rigorous research practice.

The second perspective *(conforming more to a multiple methods approach)* is silent about the integration of the respective methods, qualifying the use of multiple methods and data in a single study as sufficing for a mixed methods approach. It resonates with justifications for mixed methods research that address issues of: - comprehensiveness, multiple view points, inclusiveness and democracy, and is tailored to the 'political' needs of evaluation. The findings from this study have revealed that some evaluation researchers and practitioners conceive this latter perspective as achieving objectives embracing those of the former (i.e. richer inferences and validity of findings). However, it has been demonstrated that multiple methods research leads to multiple but scarcely richer inferences. On the contrary, multiple inferences can weaken each other when contradicting or inaccurate qualitative and quantitative findings result, especially when the complimentary function of either method is not planned a priori. This latter perspective which seems a broadening of the former conception may be eschewing or diluting the primary 'technical' purpose of using multiple methods as initially intended by the early mixed methods researchers, with the 'political' purposes *(inclusiveness, completeness, democracy)* of evaluation research taking higher priority than the 'technical' intentions.

Within the context of using mixed methods evaluation for the primary purpose of eliciting richer and more valid evaluative inferences, this study has labeled the first perspective a 'strong'

conception and the second a ‘weak’ conception of mixed methods evaluation. It is in this context that a recommendation for a reflection on the purpose and qualification of the second perspective (a multiple methods use) as a mixed methods approach is proposed. It is possible that the use of multiple methods in evaluation without the explicit intention of integration is merely a fad that may be potentially more harmful than useful to this novel evaluation approach. This fad is underscored by the findings of this study that reveal an insistence by some authors on making assertions about the benefits of the ‘mixed methods’ approach used without any evidence of this, let alone their possibility within the study designs adopted. This has implications particularly for the mixed methods purpose of ‘expansion’ and rationales of pluralism, inclusiveness and democracy, which are seemingly popular within programme evaluation practice. Mixed methods evaluation research may therefore need to reconsider some proposed purposes and designs that focus only on the mere use of multiple qualitative and quantitative methods without any requirement for the integration of the methods, data or results/findings.

# References


1. Abrams, L. S., Shannon, S. K. S., Sangalang, C. (2008). Transition services for incarcerated youth: A mixed methods evaluation study. Children and Youth Services Review, 30, 522-535.
2. Alkin, M. C. (Ed.). Evaluation roots: Tracing theorists' views and influences, Thousand Oaks, London, New Dehli: Sage.
3. Ammenwerth, E., Iller, C., Mansmann, U. (2002). Can evaluation studies benefit from triangulation? A case study. International journal of medical informatics, 70, 237-248.
4. Beach, K. D., Becker, B. J., & Kennedy, M. M. (2006). Constructing conclusions. In C. F. Conrad & R. C. Serlin (Eds.), The Sage handbook for research in education (pp. 493-509). Thousand Oaks, CA: Sage.
5. Berry, T. R., Spence, J. C., Plotnikoff, R. C., Bauman, A., McCargar, L., Witcher, C., Clark, M., Stolp, S. (2009). A mixed methods evaluation of televised health promotion advertisements targeted at older adults. Evaluation and Programme Planning, 32, 278-288.
6. Bleadsoe, K. L. and Graham, J. A. (2005). The use of multiple evaluation approaches in programme evaluation. American Journal of evaluation, 26, 3, 302-309.
7. Brazier, A., Cooke, K., Moravan, V. (2008). Using Mixed Methods for Evaluating an Integrative Approach to Cancer Care: A Case Study. Integrative Cancer Therapies, 7, 1, 5-17.
8. Brewer, J. & Hunter, A. (1989). Multimethod research: A synthesis of styles, Newsbury Park, CA: SAGE.
9. Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait-multi-method matrix. Psychological Bulletin, 56, 81-105.
10. Caracelli, V. J., Greene, J. C. (1997). Crafting Mixed-Method Evaluation Designs. In Greene, J. C. & Caracelli V. J. (Eds). Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms (pp. 19-32). San Francisco: Jossey-Bass.
11. Caracelli, V. W., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. Educational Evaluation and Policy Analysis, 15, 195-207.
12. Chen, H. T. (2004). Practical programme evaluation: Assessing and improving planning, implementation, and effectiveness. Newbury Park, CA: Sage.
13. Chen, H. T. (2006). A theory-driven evaluation perspective on mixed-methods research. Research in the schools, 13, 1, 75-83.
14. Chen, H. T. and Rossi, P. H. (1983). Evaluating With Sense: The Theory-Driven Approach. Evaluation Review, 7, 3, 283-302.

15. Collins, K. M. T., Onwuegbuzie, A. J., & Sutton, I. L. (2006). A model incorporating the rationale and purpose for conducting mixed methods research in special education and beyond. Learning Disabilities: A Contemporary Journal, 4, 67-100.
16. Cook, D.T., (2005), 'Causal Generalization: How Campbell and Cronbach Influenced My Theoretical Thinking on This Topic, Including Shadish, Cook, and Campbell' In Alkin, M.C. Evaluation Roots (2005), Thousand Oaks, CA: Sage Publication
17. Cook, T. D. (1985), "Post-Positivist Critical Multiplism," in Shotland, R. L. and Mark, M. M. (Eds.). Social Science and Social Policy. Beverly Hills: Sage, pp. 21-62.
18. Cook, T. D., Shadish, W. R. (1986), "Programme evaluation: The worldly science", pp.193 - 232.
19. Creswell J. W. and Tashakkori, A. (2007). Editorial: Differing Perspectives on Mixed Methods Research. Journal of Mixed Methods Research, 1, 4, 303-308.
20. Creswell, J. W. (1995). Research Design: Qualitative and quantitative approaches. Thousand Oaks, CA. Sage
21. Creswell, J. W. and Clark, V. L. P. (2007). Designing and Conducting Mixed Methods Research. Thousand Oaks, California. Sage Publications Ltd.
22. Creswell, J.W., Clark, V. L. P., Gutmann, M. L., Hanson, W. E.. (2003). Advanced Mixed Methods Research designs. In A. Tashakkori & C. Teddlie (Eds.), Handbook of mixed methods in social and behavioral research (pp. 209-240). Thousand Oaks, CA: Sage.
23. Cronbach, L. (1982). Designing evaluations of educational and social programmes. San Francisco: Jossey-Bass.
24. Davidson, J. (2004). Evaluation methodology basics: the nuts and bolts of sound evaluation. Sage.
25. Dellinger, A.B., & Leech, N.L. (2007). Towards a unified validation framework in mixed methods research. Journal of Mixed Methods Research, 1, 4, 309-332
26. Denzin, N. K. (1978). The research act: A theoretical introduction to sociological methods. New York: Praeger.
27. Denzin, N. K., and Lincoln, Y. S. (1994). Introduction: Entering the field of qualitative research. In Denzin, N.K., Lincoln, Y. S. (Eds.). Handbook of Qualitative Research (pp. 1-18). Thousand Oaks, CA: Sage.
28. Donaldson, S. I. and Lipsey, M. W. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In Shaw, I, Greene, J. C, Mark M. M. Sage handbook of evaluation. (Pp. 56-75). London, Southern Oaks, New Dehli: Sage.

29. Eisner, E. (1994). The educational imagination: On the design and evaluation of educational programmes (3rd ed.). New York: Macmillan.
30. Eisner, E. (1998). The enlightened eye: qualitative inquiry and the enhancement of educational practice. Upper Saddle river, NJ: Merrill.
31. Ferketich, S. L., Figueredo, A. J., and Knapp, T. R. (1991). The Multitrait-Multimethod Approach to Construct Validity. Research in Nursing & Health, 14, 315-320.
32. Festinger, I., Riecken. H. W, and Schachter, S. (1956). When prophecy fails. Minneapolis: University of Minnesota press.
33. Fossey, E., Harvey, C., McDermott, F., Davidson, L. (2002). Understanding and evaluating qualitative research. Australian and New Zealand journal of Psychiatry, 36, 717-732.
34. Giddings, L. S. (2006). Mixed-methods research: Positivism dressed in drag? Journal of research in nursing, 11, 3,195-203.
35. Greene, J. C. (1997). Advancing Mixed-Methods Evaluation. The Evaluation exchange, 3, 1.
36. Greene, J. C. (2006). Toward a methodology of mixed methods social inquiry. Research in the Schools, 13, 1, 93-99.
37. Greene, J. C. (2007). Mixed Methods in Social inquiry (1st ed.). San Francisco, CA: Jossey-Bass
38. Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? Journal of mixed methods research, 2, 1, 7-22.
39. Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issues in mixed-method evaluation. In J. C. Greene, & V. J. Caracelli (Eds), Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms (pp. 5-18). San Francisco: Jossey-Bass.
40. Greene, J. C., & Caracelli, V. J. (2003). Making paradigmatic sense of mixed methods practice. In A. Tashakkori & C. Teddlie (Eds.), Handbook of mixed methods in social and behavioral research (pp. 91-110). Thousand Oaks, CA: Sage.
41. Greene, J. C., Caracelli V. J., Graham, W. F. (1989). Toward a conceptual framework for Mixed-Methods Evaluation Designs. Education Evaluation and policy analysis, 11, 3, 255-274
42. Greene, J., Benjamin, L., Goodyear, L. (2001). The merits of mixing methods in evaluation. Evaluation, 7, 1, 25-44. London, Thousand Oaks and New Dehli: Sage.
43. Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. Newbury Park, CA: Sage.

44. Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (pp. 105-117). Thousand Oaks, CA: Sage.
45. Guba, E. G., & Lincoln,Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences.
46. House, E. R. (2004). Intellectual history in evaluation. In Alkin, M. C. (Ed.). Evaluation roots: Tracing theorists' views and influences, (pp. 218-224), Thousand Oaks, London, New Dehli: Sage.
47. Howe, K. (2003). Closing methodological divides: Toward democratic educational research. Dordrecht, Netherlands: Kluwer.
48. Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. Educational Researcher, 17, 8, 10-16.
49. Huberman, M. (1987). How well does education research really travel? Education researcher, 16, 1, 5-13.
50. Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. Administrative Science Quarterly, 24, 602-611.
51. Johnson, R. B. and Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. Educational researcher, 33, 7, 14-26.
52. Johnson, R. B. and Onwuegbuzie, A. J. and Turner, L. A. (2007). Toward a definition of mixed methods research. Journal of mixed methods research, 1, 2, 112-113.
53. Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2005). Mixed methods research: Is there a criterion of demarcation? Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
54. Kuhn, T. (1970). Postscript—1969. In T. Kuhn, The structure of scientific revolutions (2nd ed., pp. 174-210). Chicago: University of Chicago Press.
55. Kuhn, T. (1974). The essential tension: Selected studies in scientific tradition and change. Chicago: University of Chicago Press.
56. Kuhn, Thomas S. (1962). The Structure of Scientific Revolutions, Chicago: University of Chicago Press
57. Lalithamma, M. S. and Khosravi, M. No date. INADEQUACIES OF SIGNIFICANCE TESTS IN EDUCATIONAL RESEARCH. Available online: http://www.aiaer.net/ejournal/vol21109/4.%20Latchanna%20&%20Masoomeh.pdf
58. Leech, N. L. and Onwuegbuzie, A. J. (2007). A typology of mixed methods research designs. Quality and Quantity, 43, 265-275.

59. Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Beverly Hills, CA: Sage.
60. Lindsay, A. C. (2002). Integrating Quantitative and Qualitative Methods to Assess the Impact of Child Survival Programmes in Developing Countries: The Case of a Programme Evaluation in Ceara, Northeast Brazil. Health education behavior, 29, 5, 570-584.
61. Madey, D. L. (1982). Some benefits of integrating qualitative and quantitative methods in programme evaluation with illustrations. Educational evaluation and policy analysis, 4, 2, 223-236.
62. Madison, A. (2002). BRIDGING EVALUATION THEORY AND PRACTICE. Lectures presented at the African Evaluation Association (AfrEA) Conference. Available online: http://www.afrea.org/documents/Document.cfm?docID=131
63. Marsh, H. W. and Ellis, L. A. (2009). Addressing the Challenges Faced by Early Adolescents: A Mixed-Method Evaluation of the Benefits of Peer Support. American journal of community psychology, 44, 54-75.
64. Mathison, S. (1988). Why triangulate? Educational researcher, 17, 2, 13-17.
65. Maxwell, J. A. (2004). Using Qualitative Methods for Causal Explanation. Field methods, 16, 3, 243-264.
66. McConney, A., Rudd, A., & Ayres, R. (2002). Getting to the bottom line: A method for synthesizing findings within mixed-method programme evaluations. American Journal of Evaluation, 23, 2, 121-140.
67. Mertens, D. M. (2007). Transformative Paradigm: Mixed Methods and Social Justice. Journal of Mixed Methods Research, 1, 3, 212-225
68. Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50, 9, 741-749.
69. Miller, S. I. and Fredricks, M. (2006). Mixed-methods and evaluation research: trends and issues. Qualitative Health Research, 16, 567-579.
70. Mohr, L. B. (1982). Explaining organizational behavior. San Francisco: Jossey-Bass.
71. Mohr, L. B. (1996). The causes of human behavior: Implications for theory and method in the social sciences. Ann Arbor: University of Michigan Press.
72. Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications to health research, Qualitative health research, 8, 3, 362-376.
73. Morgan, D. L. (2007). Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. Journal of Mixed Methods Research, 1, 1, 48-76.

74. Morgan, J. C. and Konrad, T. R. 2008. A Mixed-Method Evaluation of a Workforce Development Intervention for Nursing assistants in nursing homes: A Case of WIN A STEP UP. The Gerontolist, 48, 1, 71-79.
75. Morse, J.M. (1991) Designing funded qualitative research. In Denzin, N.K., Lincoln, Y.S. (eds) Handbook of Qualitative Research. Thousand Oaks, CA: Sage, pp. 220–235.
76. Mouton, J. (2001). How to succeed in your master's & doctoral studies: A South African guide and resource book. Pretoria: Van Schaik.
77. Neill, J. 2008. Why use Effect Sizes instead of Significance Testing in Programme Evaluation? Available online: http://wilderdom.com/research/effectsizes.html
78. Onwuegbuzie, A. J. & Leech, N. L. (2006). Linking research questions to mixed methods data analysis procedures. The qualitative Report, 11, 3, 474-498.
79. Onwuegbuzie, A. J. (2000). Positivists, Post-positivists, Post-culturalists and Post-modernists. Why can't we all get along? Towards a framework for unifying research paradigms. Annual meeting for the advancement of educational research, Ponte Vedra, FL. November 18, 2000.
80. Onwuegbuzie, A. J., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in social science research. The Qualitative Report, 12, 2, 281-316.
81. Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. Research in the Schools, 13, 1, 48-63.
82. Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), Handbook of mixed methods in social and behavioral research (pp. 351-383). Thousand Oaks, CA: Sage.
83. Patton, M. Q. (1990). Qualitative evaluation and research methods (2nd ed.). Newbury Park, CA: Sage.
84. Perry, J. C., DeWine, D. B., Duffy, R. D., Vance, K. S. (2007). The Academic Self-Efficacy of Urban Youth: A Mixed-Methods Study of a School-to-Work Programme. Journal of Career Development. 34, 2, 103-126.
85. Phillips, D. L (1975). Paradigms and Incommensurability. Source: Theory and Society, Vol. 2, No. 1, pp. 37-61. Springer
86. Rallis, S. and Rossman, G. (2003). "Mixed methods in evaluation contexts: A pragmatic framework", in A. Tashakkori, C. Teddlie (Eds), Sage, Thousand Oaks, CA, pp.491 - 512.
87. Reichardt, C. S. & Cook, T. D. (1979). Beyond qualitative versus quantitative methods. In T. D. Cook and C. S. Reichardt (Eds.), Qualitative and quantitative methods in evaluation research (pp. 7-32). Newsbury park, CA: Sage.

88. Riggin, L. J. (1997). Advances in mixed-method evaluation: A synthesis and comment. In Greene, J. C. & Caracelli, V. J. (Eds). Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms (pp. 87-94). San Francisco: Jossey-Bass.
89. Roethlisberger, F. J. and Dickson, W. J. (1939). Management and the worker. Cambridge, MA: Harvard University press.
90. Rossi, P. H., Howard, M. W. L., Freeman, E. (2004). Evaluation: A systematic Approach. Thousand Oaks, London, New Dehli, Sage Publications.
91. Rossman, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. Evaluation Review, 9, 627-643
92. Sasaki, R. (2005). Generalizability issue: how the evaluation theorists have discussed. EVAL6010 (Interdisciplinary seminar on evaluation).
93. Schmidt,F.L.,& Hunter, J. E. (1995) The impact of data-analysis methods on cumulative research knowledge: Statistical significance testing, confidence intervals, and meta-analyses. Evaluation and the Health Professions 18, 408–427.
94. Schwandt, T. (2006). Opposition redirected. International Journal of Qualitative Studies in Education, 19, 803-810.
95. Schwandt, T. A. (2000). Three epistemological stances for qualitative inquiry. In Denzin, N. K. & Lincoln, Y. S. (Eds.), Handbook of qualitative research (2nd ed., pp. 189-213). Thousand Oaks, CA: Sage.
96. Scriven, M. (1966). Value claims in the social sciences (Social science education consortium publication No. 123). Lafayette, IN: Purdue University.
97. Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), Perspectives of curriculum evaluation (pp. 39-83). Chicago: Rand McNally.
98. Scriven, M. (1972a). The methodology of evaluation. In C. H. Weiss (Ed.), Evaluating action programmes: Readings in social action and education (pp. 123-136). Boston: Allyn & Bacon.
99. Scriven, M. (1980). The logic of evaluation. Inverness, CA: Edgepress.
100. Scriven, M. (1983b). The evaluation taboo. In E. R. house (Ed.), Philosophy of evaluation (pp. 75-82), San Francisco: Jossey-Bass.
101. Scriven, M. (1991). Evaluation thesaurus (4th Ed.). Newbury Park, CA: Sage.
102. Scriven, M. (1994). The final synthesis. Evaluation Practice, 15, 3, 367-382.
103. Scriven, M. (2003). Evaluation theory and meta-theory. In Kellaghan, T. & Kellaghan, T and Stufflebeam, D. L. (Eds.). International Handbook of Educational Evaluation, (pp. 15-30). Dordrecht: Kluwer Academic

104. Scriven, M. (2005). The logic and methodology of checklists. Available online: www.wmich.edu/evalctr/checklists/
105. Scriven, M. S. (1983a). Evaluation ideologies. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.). Evaluation models: View points on educational and human services evaluation (pp. 229-260). Boston: Kluwer-Nijhoff.
106. Shadish W. R, Cook T. D, & Leviton L. C. (1991). Foundations of programme evaluation: Theories of practice. Newbury Park, London, New Dehli: Sage.
107. Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., and Sherif, C. W. (1961). Intergroup conflict and cooperation: The Robber's cave experiment. Norman: University of Oklahoma, Institute of Intergroup Relations.
108. Sieber, S. D. (1973). The integration of fieldwork and survey methods. American Journal of Sociology, 73, 1335-1359.
109. Smith, M. L. (1994). Qualitative plus/versus quantitative: The last word. In C. S. Reichardt & S. F. Rallis (Eds.), The quantitative-qualitative debate: New perspectives (pp. 37-44). San Francisco: Jossey-Bass.
110. Stake, R. & Schwandt, T. (2006). On discerning quality in evaluation. In Shaw, I.F, Greene, J.C, & Mark, M. M (Eds.). Handbook of Programme Evaluation. London: Sage.
111. Stake, R. E. (1995). The Art of Case Study Research. Thousand Oaks, CA: Sage Publication
112. Stake, R., Migotsky, C., Davis, R., Cisneros, E.J., DePaul, G., Dunbar, Jr., C., Farmer, R., Feltovich, J., Johnson, E., Williams, B., & Chaves, I. (1997). The evolving syntheses of programme value. Evaluation Practice, 18, 2, 89-103.
113. Stake. R. (2004). Stake and responsive evaluation. In Alkin, M. C. (Ed.). Evaluation roots: Tracing theorists' views and influences, (pp. 203-217), Thousand Oaks, London, New Dehli: Sage.
114. Stern, P. N. (1994). Eroding grounded theory. In M. J. Morse (Ed.), Critical issues in qualitative research methods (pp. 214-215). Thousand Oaks, CA: Sage.
115. Symonds, J. E. and Gorard, S. The Death of Mixed Methods: Research Labels and their Casualties. The British Educational Research Association Annual Conference, Heriot Watt University, Edinburgh, September 3-6, 2009
116. Tashakkori, A. & Teddlie, C. (2003), Handbook of mixed methods in social and behavioral research, Thousand Oaks, CA: Sage.
117. Tashakkori, A. & Teddlie, C. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.),

Handbook of mixed methods in social and behavioral research (pp. 3-49). Thousand Oaks, CA: Sage.

118. Tashakkori, A. and Creswell, J. W. (2007). Editorial: Exploring the nature of research questions in Mixed Methods Research. Journal of Mixed Methods Research, 1, 3, 207-211.

119. Tashakkori, A. and Teddlie, C. (1998) Mixed Methodology: Combining Qualitative and Quantitative Approaches. Thousand Oaks, CA: Sage.

120. Teddlie, C. & Yu, F. (2007). Mixed Methods sampling: A typology with examples. Journal of Mixed Methods Research, 1, 1, 77-100.

121. Teti, M., Bowleg, L., Cole, R., Lloyd, L., Rubinstein, S., Spencer, S., Aaron, A., Ricksecker, A., Berhance, Z., Gold, M. (2009). A Mixed Methods Evaluation of the Effect of the Protect and Respect Intervention on the Condom Use and Disclosure Practices of Women Living with HIV/AIDS. AIDS Behav, 11, S106–S116.

122. Trokim, W. M. K. (2006). Internal validity. Available online: http://www.socialresearchmethods.net/kb/intval.php

123. Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). Unobtrusive measures. Chicago: Rand McNally.