



The Indispensability of Utilitarian Evaluation Design in the Reformation of Nigerian Education

Peter James Kpolovieⁱ
University of Port Harcourt, Nigeria
drkpolovie@yahoo.com

&

Nwachukwu Prince Ololubeⁱⁱ
University of Education, Port Harcourt, Nigeria
ololubepriince@yahoo.com

Abstract

This article avers that the application of a utilitarian evaluation design adopted in pure and unbiased evaluation research to produce the very best decision alternatives about a program that it is hoped will be of greatest practical utility and propriety to all stakeholders of the program and society in general, is indispensable in the genuine reformation of Nigerian education. Utilitarian evaluation design maximizes the satisfaction that society can derive from the program under investigation. This research design uses cost-benefit analyses that are utilitarian, egalitarian, scholarly, ethical, and libertarian in the evaluation of an education program to achieve the best outcomes for all stakeholders. Without thorough and due evaluation, no program has been, and perhaps none will ever be, successful. For any component of the Nigerian education system to be successful by international standards, it must be regularly subject to utilitarian evaluation design.

Key words: Utilitarian evaluation design, Program evaluation, Education reform in Nigeria, Utilitarian evaluation execution guidelines, Evaluation research, Evaluation design, Evaluation approaches, Pseudo-evaluation.

Reference to this paper should be made as follows:

Kpolovie, P. J., & Ololube, N. P. (2013). The Indispensability of Utilitarian Evaluation Design in the Reformation of Nigerian Education. *International Journal of Scientific Research in Education*, 6(2), 159-178. Retrieved [DATE] from <http://www.ij sre.com>

INTRODUCTION

An evaluation research design, or evaluation design, is a detailed description of the processes and procedures for arriving at, and disseminating, the best data-based decision alternatives about the merit and worth of a program under specific circumstances. It is the exhaustive explanation of the approaches and methods adopted for gathering, analysing and interpreting the data that serve as the basis for drawing decision alternatives for maximization of the merit and worth of a program (The World Bank Group, 2012). Evaluation design portrays in clear terms, the procedure(s) employed in assessing the utility of a program; explicitly showing the resources, presuppositions, potentialities, consequences, limitations, main cues used, major purpose(s) served, typically adopted methods, representative samples studied, questions that characterize the investigation, and the main considerations for determining when it is appropriate to adopt the design (Kpolovie, 2010). Evaluation designs or approaches are conceptually distinct ways of thinking about, structuring and conducting evaluation efforts for the making of unique contributions to the resolution of important problems and the refinement of programs for the exceptional accomplishment of both predetermined and unintended goals.

Program evaluation is an essential component of every meaningful reform or intervention in the education sector. Evaluation involves the systematic collection, analysis, and use of information to answer basic questions about the overall effectiveness of a program or about specific services or activities implemented through the program. According to James Bell Associates (2009), the term *systematic* denotes the use of structured and consistent methods for collecting and analysing information. Any systematic data collection and analysis effort thus requires the selection of an overarching evaluation research design to guide data collection activities and ensure that they are implemented in a valid, reliable and coherent manner. The various evaluation designs on which the reform of education programs must be based are examined here. Emphatically, education reform that is indeed progressive cannot be done without the execution of evaluation research that uses the most appropriate evaluation design as its starting point.

An evaluation research design refers to the overarching methodological framework that guides an evaluation effort. It is the conceptual lens through which the evaluation is viewed and implemented. Evaluation research design “provides the glue that holds the research project together. Evaluation design is used to structure the research, to show how all of the major parts of the research project ... work together to ... address the central research questions” (James Bell Associates, 2009). Trochim (2006) deemed evaluation design to be either “the systematic assessment of the worth or merit of some object” or “the systematic acquisition and assessment of information to provide useful feedback about some object”. The latter definition was offered because there are many types of evaluations that do not *necessarily* result in an assessment of worth or merit – descriptive studies, implementation analyses, and formative evaluations, to name a few. The definition, in other words, emphasizes the information-processing and feedback functions of evaluation. Both pieces of Trochim’s definition converge around the fact that evaluation is a *systematic* endeavour and both use the deliberately ambiguous term 'object' which could refer to a program, policy, technology, person, need, activity, and so on. The latter definition emphasizes *acquiring and assessing information* rather than *assessing worth or merit* in that all evaluation work involves collecting and sifting through data, making judgments about the validity of the information and the inferences derived from it, regardless of whether an assessment of worth or merit results.

EVALUATION APPROACHES

Evaluation approaches are used interchangeably with evaluation designs in this chapter, except where otherwise specified. Evaluation designs are the way in which the evaluation ingredients – approach, purposes and methods – are put together into the final evaluation in an attempt to answer a set of evaluation questions. If there is to be an outcome evaluation within the evaluation, the design will specify which of the many types of outcome design are to be used. Evaluation designs or approaches are the overall ways of conceptualizing evaluation, including philosophical and value orientations to the task of undertaking evaluations. Evaluation designs, in the widest sense, are the overall evaluation approach, the mix of input, formative, process and impact evaluation, and the methods and analysis techniques to be used, together with the timing of when various types of information will be collected. Evaluation designs are useful in thinking about the philosophical and other postures which can be adopted in regard to evaluation. Evaluation designs can be seen as the way in which the other evaluation ingredients - *types (purposes), methods and analysis techniques* are put together in order to answer evaluation questions (Knol, 2012).

There are a considerable number of evaluation designs. The reform of education programs in Nigeria, as in any other nation, must be clearly based on at least one of these. The various evaluation designs that are adopted in assessing the merit of educational programs and which, in turn, determine education reforms can be broadly classified into four designs with unique modus operandi and orientations. These are:

- Social mission/advocacy-oriented evaluation designs;
- Program improvement/accountability-oriented evaluation designs;
- Questions/methods-oriented evaluation designs; and
- Pseudo-evaluation designs.

Social Mission/Advocacy-Oriented Evaluation Designs

Social mission/advocacy-oriented evaluation designs are employed so as to make a difference in societal advancement through education. They seek to ensure that all segments of society have equal access to educational and other social opportunities and services. They bear an affirmative action bent toward ensuring preferential treatment through program evaluation to all, and the disadvantaged in particular. Knowledge, they say, is power. Social mission/advocacy oriented evaluation designs use program evaluation to increase common social benefit by empowering the disenfranchised majority *ab initio*. They employ the perspectives of stakeholders as well as of experts in characterizing, investigating, and judging programs for the generation and implementation of decision alternatives that directly improve society.

These designs favour a constructivist orientation and the use of qualitative as well as quantitative methods. Sometimes, they eschew the possibility of finding right or best answers and reflect the philosophy of postmodernism, emphasizing cultural pluralism, moral relativity, and multiple realities. They provide for democratically engaging stakeholders in obtaining and interpreting findings. Social mission/advocacy oriented evaluation designs meet the standards of sound evaluation while concentrating heavily on a social mission. By giving stakeholders the authority to make key evaluation decisions, related especially to the interpretation and release of findings, evaluators empower these individuals to use evaluation to their best advantage for the good of all or at least of a significant majority.

Though intent on serving the underprivileged, empowering the disenfranchised, and/or righting educational and/or social injustices, evaluators using these designs should never compromise the independent and impartial perspectives needed to produce only highly valid and reliable findings. Evaluators must thus guard against compromising the integrity of the evaluation process to achieve social objectives. . It is also important for the evaluator using these designs to subject them to meta-evaluations that are exceptionally anchored on best standards for sound evaluation.

Social mission/advocacy oriented evaluation designs are highly recommended because they are strongly oriented to the democratic principles of equity and fairness and employment of practical procedures for involving the full range of stakeholders to attain utilitarian objectives. The evaluation designs that constitute social mission/advocacy oriented evaluation designs are:

- Utilitarian evaluation design;
- Utilization-focused evaluation design;
- Responsive evaluation design;
- Deliberative democratic evaluation design; and
- Constructivist evaluation design.

Program Improvement/Accountability-Oriented Evaluation Designs

Program improvement/accountability-oriented evaluation designs stress the need to fully assess a program's merit and worth for optimum enhancement of the program. They consider the full range of questions and criteria required to assess the true value of a program, examining stakeholders' needs as well as the technical and economic criteria pertinent to judging a program's inputs, operations and outputs. They assess all relevant outcomes, not just those outlined as intended objectives of the program. Often, they are objectivist and assume an underlying reality in seeking

definitive, unequivocal answers to the evaluation questions. Program improvement/accountability oriented evaluation designs typically use multiple qualitative and quantitative assessment methods to provide crosschecks on findings. They emphasize improvement by delivering program decisions, providing consumers with assessments of optional programs and services, and helping consumers to gain unquestionable assurances that the program in question is professionally sound and effective *vis-à-vis* other similar or competing programs. Specific evaluation designs that constitute program improvement/accountability oriented evaluation designs are:

- Consumer/client centred evaluation design; and
- Accreditation/certification evaluation design.

Questions/Methods-Oriented Evaluation Designs

Questions/methods-oriented evaluation designs use specific preferred method(s) to address predetermined limited questions that often yield answers which are insufficient to assess the entire merit and worth of the program under investigation. These designs do not consider the adequacy and appropriateness of the adopted questions and methodologies in the systematic assessment of a program's merit and worth. Generally, questions/methods-oriented evaluation designs emphasize technical quality and posit that it is usually better to answer a few pointed questions well than to attempt a broad assessment of merit and worth. Most times, the focus of questions/methods evaluation designs is either too narrow or is only tangential or peripheral to questions of merit and worth, although in a few cases, these designs do happen to provide evidence that fully assesses an object's merit and worth.

Potter (2006) identifies and describes three broad paradigms within program evaluation. The first, and probably most common, is the positivist approach, in which evaluation can only occur where there are "objective", observable and measurable aspects of a program, requiring predominantly quantitative evidence. The positivist approach includes evaluation dimensions such as needs assessment, assessment of program theory, assessment of program process, impact assessment and efficiency assessment plus a number of those that constitute questions/methods-oriented evaluation designs (Rossi, Lipsey and Freeman, 2004). A recent study conducted by the Public Policy Institute of California, "Evaluating Academic Programs in California's Community Colleges", in which the evaluators examine measurable activities (i.e. enrolment data) and conduct quantitative assessments like factor analysis offers an example of the positivist approach. While the designs are typically labelled as evaluations, they may or may not meet the stringent requirements of a sound evaluation. Consequently, questions/methods evaluation designs are often termed quasi-evaluation designs (Stufflebeam, Madaus & Kellaghan, 2002). Quasi-evaluation designs have legitimate uses, since they can investigate important, though narrow, questions about a program or an object. The main caution is that these types of studies may not be uncritically equated to pure evaluation. The evaluation designs that constitute questions/methods oriented evaluation designs are:

- Case study evaluation design
- Benefit-cost analysis evaluation design
- Criticism/connoisseurship evaluation design
- Mixed-method evaluation design
- Program theory-based evaluation design
- Clarification hearing evaluation design
- Management information systems evaluation design
- Experimental studies evaluation design
- Performance testing evaluation design
- Outcomes as value-added evaluation design
- Payment-by-results evaluation design

Pseudo-Evaluation Designs

Pseudo-evaluation designs either do not qualify as evaluation or represent questionable practices that do not meet the state of soundness that program evaluation demands. Pseudo-evaluation designs either do not validly assess the program or assess the program but shade, selectively release, or even falsify the findings. Pseudo-evaluation designs fail to produce and report, to all right-to-know audiences, valid assessments of merit and worth of the object under study.

Pseudo-evaluation designs are often adopted by those in power to either mislead audiences and constituents or to gain and maintain an unfair advantage over others, particularly the majority the program is designed to serve. When adopted by insensitive politicians, they designs can promote, support and perpetrate injustice, mislead decision making, lower confidence in evaluation services, and discredit the evaluation profession. Persons holding or seeking political authority use pseudo-evaluations, for example, to either deceptively introduce a new program of little or no merit and worth to the populace or to maintain a program that should have, on the basis of needs assessment, been terminated for lack of value and utility. With pseudo-evaluation, ill-minded politicians are able to present unwarranted claims about their achievements and/or the faults of their opponents, or hide potentially damaging information.

Education reforms that are not based on evaluation research design, but rendered to feed the greedy habits and addictions of the few in power can rightly be said to have been anchored in pseudo-evaluation. This is because pseudo-evaluation generally means the absence of evaluation (Wikimedia Foundation Inc., 2011; Stufflebeam, Madaus & Kellaghan, 2002). Politically controlled or politically-oriented evaluation design and public relations, or legitimacy evaluation studies, are based on an objectivist epistemology from an elite perspective. Although these approaches seek to misrepresent value interpretations about some object, they go about it in a manner that defeats the very essence of program evaluation. Information obtained through politically controlled studies is released or withheld to meet the special interests of the holder, usually the politician. Legitimacy evaluation design thus paints a positive image of an object regardless of the actual situation. It is not an acceptable evaluation practice (Frisbie, 1986), although it is the only one that education reform in Nigeria has ever been based on as clarified in Kpolovie (2012).

Utilitarian Evaluation Design

Of the four evaluation approaches, this paper recommends use of the social mission/advocacy-oriented evaluation design. Of the five evaluation designs that constitute social mission/advocacy evaluation, the utilitarian evaluation design is preferred (Kpolovie, 2012; Joint Committee on Standards for Educational Evaluation, 1998), accounting for why this design has been chosen as the focus of this chapter. The utilitarian evaluation research design adopts ethical, moral, justice and fairness principles in studying the inputs, operations or process and outputs of a given program to fashion ways forward that will better ensure attainment of the goals of the program based on a societal needs assessment. These ways forward highlight the morally appropriate actions that will ultimately maximize individual, organizational and societal value for operating the program, and may even indicate whether or not the society at hand would be better off without the particular program (Kpolovie, 2010; Wikimedia Foundation, Inc., 5 October, 2011; Ohio Department of Education, 2011). Utilitarian evaluation design is a research methodology that satisfies both scientific and ethical requirements for arriving at and passing value judgments on the way that a program is planned and executed for the actualization of related needs (Christie & Fliescher, 2011; Bennett, 2011; University of Wisconsin-Madison, 2011; Wikimedia Foundation, Inc., 2011a; Siy, 2007). It allows for an understanding of the aims and interests of the evaluators, operators of the program, sponsors of the research, members of the program, subjects of the study and members of the society from their different perspectives or standpoints (Sun Associates, 1998).

The design emphasizes the accuracy of findings by preventing sources of bias. It ensures or insists on the utility of results for all project stakeholders, feasibility of the findings in terms of their applicability and propriety in terms of full and frank disclosure, the public's rights to know, the rights of human subjects, balanced reporting, and fiscal responsibility. In situations where the research is being sponsored by a government or other agency, the evaluator enters into a written contract with the sponsor in terms of privacy, the protection of human subjects (i.e., informed consent) and freedom of information release. Such contracts contain lists of ethical responsibilities that the evaluator and sponsor must respectively adhere to. It does not allow for the sponsor's interference in any aspect of the evaluation. The design thus maximizes outcomes as the evaluation is valid, the evaluators are competent, participants are informed, consequences of the evaluation are weighed against others, and the results are disseminated to legitimate

audiences therefore allowing other professionals to examine the procedures and data for possible publication of rejoinders to misinterpretations (if any) of results. The design respects autonomy and ensures justice by reducing the power differential between the evaluator and participants, informing participants and having them volunteer, and equally treating and representing sub-groups within society in the sample (Haddad & Demsky, 1995; Porter, 2011).

Utilitarian evaluation design is pluralistic in nature as it accommodates the presumption that there are multiple end principles of justice. It is in this regard that the design broadly elicits the needs and perceptions of various interest groups with respect to the program under investigation. In other words, the design treats interest groups as several in number, each having its distinct interests, since the pluralist principle holds that there is no overriding endpoint or measure of general welfare. Thus, with utilitarian evaluation design, the interests of even disadvantaged or powerless groups are represented and accorded priority in both the evaluation and the program under study (Wikimedia Foundation, Inc., 2011a; Siy, 2007). In this way, the design is able to ensure that no one individual or group imposes its will upon others through coercion, force or illegitimate means with regard to the planning, implementation and goals of an educational program (Murtala, 2011). It was through utilitarian evaluation design, for example, that nomadic education for cattle rearers' children in Nigeria was eventually extended to cover or include migrant fishermen's children. Cattle-rearing is done mostly by members of a major ethnic group, Hausa, while fishing is done mostly by members of the minority ethnic group, Izon.

The fundamental notion of equality, that all should be seen as members of the same reference group and consequently be treated the same way in terms of the worth or usefulness of a program, is best accounted for by the utilitarian evaluation design. In the public determinations of wants on which the goals of an educational program should be based, utilitarian design dictates that the satisfaction of each person's or group's interests merits equal consideration. The design impartially settles the inevitable conflicts that emanate from allowing each individual or group to advance its interests about a program under investigation. This is because unlike policy-oriented evaluation design and legitimacy design, both of which are biased in favour of specific politicians and administrators of educational institutions, utilitarian evaluation design is impartial in its procedures and ensures fair consideration of all groups and sub-groups (Wikimedia Foundation, Inc., 2011; Harrall, Burt, Hatry & Ressler, 2011; Sanctuary Web Team, 2011; Lichtenstein, 2011; NOAA, 2011; Greene, Caracelli & Graham, 2011; The Design-Based Research Collective, 2011; Brown University, 2011).

Utilitarian evaluation design can investigate the effectiveness, efficiency and accountability of an educational program as planned and executed for the overall good of the society. The design seeks to provide data-based answers to questions like the following (House, 2009; Kpolovie, 2010; Siy, 2007; Stanford Encyclopedia of Philosophy, 2009; HKU, 2011; Luca, 2010):

- What are the needs of the society and of sub-groups in the society?
- What are the needs of the stakeholders in the program under study?
- What are the goals of the program?
- What should the goals of the program be?
- What results were intended by the program?
- What results were actually obtained by the program?
- Is there a discrepancy between the intended and actually achieved results of the program?
- Of what value are the methods adopted by the program in attaining its objectives?
- What better methods and means could the program adopt to accomplish its objectives?
- How well was the program been organized and implemented?
- What, if anything, about the program organization and implementation should be changed?
- What, if anything, should be added to the program to increase utility?
- Which parts of the program should be continued (i.e. allowed to remain without modification)?
- Should the program be allowed to exist at all?

The bedrock of utilitarian evaluation design is ensuring the adequate realization of the standard of good programs. Every good program must effectively reach and serve beneficiaries' targeted needs at a reasonable cost, and do so far better than, or at least as well as other reasonably available alternative programs. It is for this reason that utilitarian

evaluation design ensures that data-based answers are provided to the fourteen (14) major questions above; as well as the twenty-one (21) additional questions posed below (Kpolovie, 2012):

- Has an appropriate beneficiary population been determined?
- What beneficiary needs should be addressed?
- What are the available alternative ways to address these needs?
- What are their comparative merits and costs?
- Are the plans of services and participation sound?
- Is there adequate provision of facilities, materials, and equipment?
- Is the program staff sufficiently qualified and credible?
- Have appropriate roles been assigned to the different participants in the program?
- Are the participants effectively carrying out their assignments?
- Is the program working?
- Should the program be revised in any way?
- Is the program effectively reaching all targeted beneficiaries?
- Is the program meeting participants' needs?
- Have beneficiaries satisfactorily played their part?
- Is the program better than competing alternatives?
- Is the program affordable?
- Is the program sustainable?
- Is the program transportable or applicable to any other similar population?
- Is the program worth the required initial investment?
- Is the program serving the greatest good for all or at least for the significant majority of stakeholders?
- Is the program actually meeting the needs of these people?

With utilitarian evaluation design, stakeholders can better enjoy the full benefits of a program that meets their psychological, social, emotional, financial, and other needs (University of Pennsylvania, 2007). In this regard, utilitarian evaluation design is guided by a positive psychology for all stakeholders of the program. Positive psychology is the scientific study of the strengths and virtues that enable individuals and communities to thrive. This is because utilitarian evaluation promotes increased research, training, education, and the dissemination of the greatest possible knowledge about the program to every stakeholder and all those who have the right to know. A central principle of utilitarian evaluation design is such that all stakeholders of a program deserve to lead the most meaningful and fulfilling life, cultivate what is best within themselves, and to optimally enhance their experiences of love, work, and play through the program.

Utilitarian evaluation design advances three key virtues – positive individual traits, positive emotions and positive institutions – that advance societal development demands. Utilitarian evaluation design promotes positive individual traits by revealing participant strengths and virtues, such as the capacity for love and work, courage, compassion, resilience, creativity, persistence, curiosity, integrity, self-knowledge, moderation, self-control, and wisdom. The design promotes positive emotions by revealing contentment with the past, happiness in the present and feasible or practical hope for the future with regard to the program's inputs, operation and outputs. Utilitarian evaluation design advances positive institutions by revealing the characteristics that foster better communities, such as justice, responsibility, civility, parenting, nurturance, work ethic, leadership, teamwork, purpose, and tolerance. Some of the additional targets of utilitarian evaluation design are the building of a program that supports:

- Families and learning institutions that allow children or students to flourish;
- Workplaces that foster the greatest satisfaction and highest productivity;
- Communities that encourage pure civic and positive deliberative engagement;
- Program providers and sponsors who identify and nurture their clients' strengths;
- The teaching of utilitarian values; and
- Dissemination of utilitarian interventions in organizations, communities and societies.

Utilitarian evaluation design emphasizes that program evaluation should be used proactively to help improve programs as well as retroactively to judge their merit and worth in meeting the needs of program stakeholders. Utilitarian evaluation design is underpinned philosophically by an objectivist orientation to finding the best answers to context-limited questions and subscribes to the principles of a well-functioning democratic society, one based on human rights, equity, excellence, conservation, and accountability. The design practically engages stakeholders in focusing the evaluation, addressing their most important questions, providing timely and relevant information decision alternatives, and providing an authentic record of program accountability. The design achieves these by matching the evaluation to the characteristics of the program itself, and to the particular resources and constraints inherent in each evaluation context (Wikimedia Foundation, Inc., 10 November, 2011; Kpolovie, 2011; McNamara, 2011) through:

- Needs assessment;
- Monitoring and accountability;
- Quality review, program clarification and quality assurance;
- Achieving intended and even unintended utilitarian objectives; and
- Establishing the impact of the program on stakeholders and society.

The underlying purpose of utilitarian evaluation design is not to prove the merit and worth of a program, but to maximally improve the lives of stakeholders through the rendering of an exceptionally enhanced program. For this reason, utilitarian evaluation design provides a value base for enacting and being accountable for decisions that result in developing, delivering, and making informed use of cost-effective services. Evaluator(s) must thus interact with truly representative members of the stakeholder groups, discern their questions, and supply them with relevant, timely, efficient, and accurate information.

The design stresses assessments of stakeholders' needs, problems, questions, and opportunities; identification and assessment of competing programs; assessment of program plans; assessment of staff qualifications and performance; assessment of program facilities and materials; monitoring and assessment of process; assessment of intended and unintended short-range and long-range outcomes; and assessment of the cost-effectiveness of the program. As a result the design can provide information on the best goals and priorities, choosing from among competing services, budgeting, staffing, service utilization, guiding participation, judging progress, and recycling the program's operations to best meet the identified needs of the stakeholders.

The stakeholders include all persons and groups who must make choices related to initiating, planning, implementing, and using the program's services such as the beneficiaries (students), parents/guardians, service providers, administrators, program consultants, support personnel, policy makers, funding authorities or program providers, taxpayers and both the immediate community and wider society. Of note, the stakeholders should also include the *marketplace* or the employers of labour. Here marketplace refers to "the dynamic world of industry, business and commerce with public, large, medium and small firms as well as individual entrepreneurs as the key players; and which dictates the demand and supply of professionals, highly skilled, skilled, semi-skilled and unskilled labour for enhancement of job flows that are characterized with hiring, firing, promoting, retiring and hopping from one job to another for the ultimate purpose of ensuring exchange of worthwhile goods and services" (Kpolovie, 2012).

Utilitarian evaluation design ensures maximum persuasive utilization, direct utilization and conceptual utilization of the program evaluation by stakeholders. Perceptual utilization is the enlistment of evaluation results in an effort to objectively persuade the stakeholders to either support or oppose certain aspects of the program agenda. This use is only of interest to the evaluator if, and only if, the stakeholders are part of those who executed the evaluation. Direct utilization of evaluation occurs when the evaluation produces results that have a direct positive effect or influence on the stakeholders through improvement of the structure or process of the program, indeed, the primary purpose of utilitarian evaluation design. Conceptual utilization occurs when results of the evaluation are enthusiastically applied in positively conscientualizing stakeholders with regards to the issues that formed part of the concerns of the program evaluation. Utilitarian evaluation design helps to ensure maximum use of the program evaluation in five simple ways:

- Evaluators must understand the cognitive styles of the decision-makers;
- Evaluation results must be timely and available when needed;
- Evaluations must respect stakeholders' program commitments;
- Utilization and dissemination plans should be part of the evaluation design; and
- Evaluations should include an assessment of utilization (Wikimedia Foundations Inc., 10 November, 2011).

A number of other research methods and their specific designs may additionally be applied in the utilitarian evaluation design for a given program. These include cross-sectional surveys, the various longitudinal survey designs, action research, ex post facto research, triangulation research, case study or single-subject research, historical research, ethnographic research, instrumentation research, correlational research, quasi-experimental research, experimentation research, and needs analysis research (Kpolovie, 2010).

The success of the utilitarian evaluation design depends largely on regular interaction of the evaluator(s) with a truly representative sample of stakeholders as advisory panel to help define the evaluation questions, shape evaluation plans, review draft reports, and help disseminate findings. This is a veritable way of ensuring that due feedback from the program evaluation are regularly provided to stakeholders for the best satisfaction of their needs through optimal program improvement. Feedback, like interim reports and formative evaluation reports, is indispensable in enabling and motivating targeted beneficiaries, program staff and other stakeholders to acquire data-based knowledge about the merits and worth of the program, the quality of their own participation, and the rapid improvement of stakeholders and society through the program. Such reports are of prime importance in the evaluator's compilation of a comprehensive summative evaluation report that provides program personnel and stakeholders with findings for the composition of their own accountability reports, which are further revolved for additional program improvement and the greatest benefits for stakeholders.

In utilitarian evaluation, stakeholders' full involvement in program evaluation is a key principle of the process of individual and societal or collective positive change. This is because the evaluation of any given program can best guarantee the desired positive change in the target group's behaviour only if the group has been involved in planning, implementing, monitoring, and empirically judging the program and its evaluation. When stakeholders are involved throughout the evaluation process, the evaluator(s) thus lays a foundation for not only helping stakeholders to understand and value the evaluation process and the program more rationally, but to use evaluation continuously and systematically by apply the findings to ensure maximal program improvement for the greatest good of all the stakeholders – the basic philosophy of utilitarianism and utilitarian evaluation design.

UTILITARIAN EVALUATION EXECUTION GUIDELINES

Utilitarian evaluation can be conducted at the various stages of a program's lifecycle. At each of the stages, different questions demand to be answered, and correspondingly different types of evaluation are applied. To meet answer these diverse questions, five forms of assessment, three conditions for causation, three requirements of a good sample, and three qualities of measuring instruments are discussed below (Rossi, Lipsey & Freeman, 2004; Kpolovie, 2010; Kpolovie, 2011; Kpolovie, 2002).

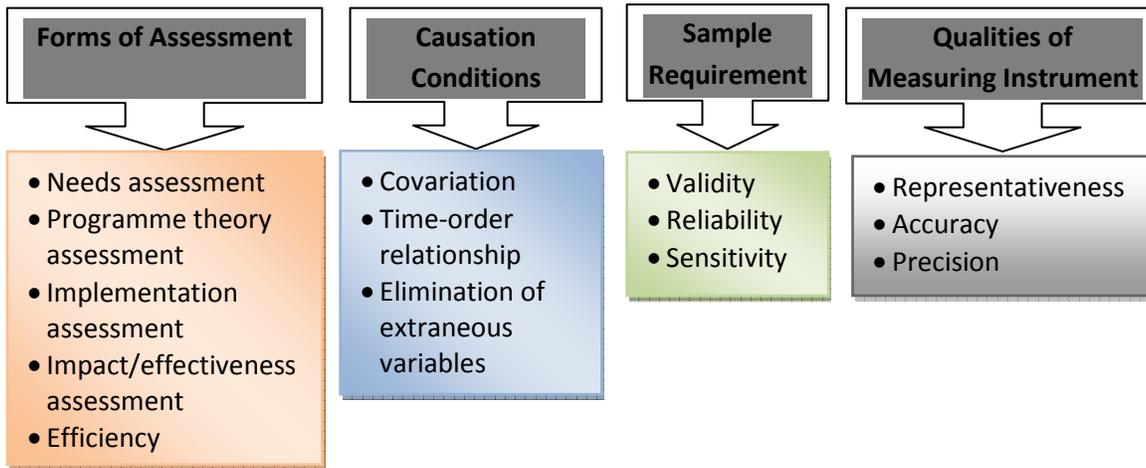


Figure 1: Utilitarian evaluation execution guidelines

FORMS OF ASSESSMENT

The five forms of assessment are as follows:

Needs Assessment

Evaluators must carry out needs assessments. Needs analysis is a crucial step in evaluating programs because the effectiveness of a program cannot be assessed unless the initial problem or need is known. A needs assessment examines a program's targeted population to see whether the need as conceptualized in the program actually exists in the population; whether it is, in fact, a problem, and if so, how it might best be dealt with. This includes identifying and diagnosing: the problem the program is trying to address, who or what is affected by the problem, how widespread the problem is, and the measurable effects caused by the problem. In a housing program aimed at mitigating homelessness, for example, a program evaluator may wish to assess how many people are homeless in a given geographic area as well as their demographic information. Rossi, Lipsey and Freeman (2004) cautioned against undertaking an intervention without properly assessing the need for it, as non-existent or misconceived needs can result in a great deal of wasted funds.

Needs assessment involves the processes or methods used to describe and diagnose existing social needs in a population (Rouda & Kusy 1995; Centers for Disease Control and Prevention 1999; Barbazette, J. 2006). This is essential for utilitarian, and other, evaluators because a program's effectiveness cannot be assessed without an identification of the initial problems or needs. Programs that fail to complete needs assessments can fall victim to the illusion that they have eradicated the problems/needs when in fact there were no needs, or worse, when the program itself has caused the needs/problems. Needs assessment involves research and regular consultation with community stakeholders and those that will benefit from the program. Needs assessment is thus a bottom-up approach as potential problems are realized earliest when the community is directly involved in identifying needs and potential barriers.

The important tasks of a program evaluator are thus to: (1) construct a precise definition of what the problem is. Evaluators need to first identify the problem/need. (2) Assess the extent of the problem. Having clearly identified what the problem is, evaluators need to then assess the extent of the problem. They need to answer the questions of 'where' and 'how big'. Undoubtedly, pointing out that a problem exists is much easier than having to specify where it is located and how rife or endemic it is.

In an example by Rossi, Lipsey & Freeman (2004), one person identifying battered children may be enough evidence to persuade audiences that child abuse exists. Indicating how many children are affected and where the abuse is taking place (both geographically and socially) requires knowledge about the abused children, the perpetrators and the impact of the problem throughout the region in question. This is a particularly challenging example, as child abuse is not a public behaviour and estimates of the rates of private behaviours are usually not reliable because of unreported

cases. In this case, evaluators would have to use data from several sources and apply different approaches in order to estimate incidence rates.

A thorough needs assessment must also answer the questions of 'how' and 'what'. The 'how' question requires evaluators to determine how the need will best be addressed. Having identified the need and familiarized themselves with the community, evaluators should conduct a performance analysis to identify whether the proposed program plan will be able to eliminate the need. The 'what' question requires evaluators to conduct a task analysis to determine the best way of solving the problem or meeting the need. This might include assessing whether the job performance standards should be set by an organization or whether certain government regulations should be considered when executing the program or undertaking the evaluation task.

Lastly, evaluators must define and identify the target of interventions and accurately describe the nature of the service needs of that population. It is important to know what/who the target population is/are – individuals, groups, communities, etc. There are three units of the population: *population at risk*, *population in need* and *population in demand*. Population at risk are those with a significant probability of developing the risk (the at risk population for birth control programs, for example, are women of child bearing age). The population in need are those with a condition that the program seeks to address (the population in need for a program that aims to provide ARV's to HIV positive people, for example, are those that are HIV positive). The population in demand are those in need who admit that they are in need and are willing to take part in what the program has to offer (not all HIV positive people, for example, will be willing to take ARV's) (Van-der-Riet, 2009). Specifying what/who the target populations are will assist in establishing appropriate boundaries so that interventions can easily assist the target population and be applied efficiently.

There are four steps in conducting a needs assessment (Centers for Disease Control and Prevention 1999):

- Evaluators need to assess the current situation through a 'gap' analysis to identify the desired situation and the actual current situation. The difference or 'gap' between the two situations will help to identify the need, purpose and aims of the program.
- Proffer a number of interventions that could potentially address the need. These must now be examined in relation to their significance to the program's goals and constraints and prioritized accordingly. This should be done in consideration of the following: cost effectiveness (consider the budget of the program, assess cost/benefit ratio), and executive pressure (whether top management expects the solutions, and whether many key people are involved).
- Identify specific problem areas within the need that are to be addressed and assess the skills of those carrying out the interventions or implementing the program.
- Identify the possible consequences of the solutions and growth opportunities as a result of the interventions.

PROGRAM THEORY ASSESSMENT

Program theory, also called a logic model or impact pathway (Miller & Salkind, 2002), is an assumption, implicit in the way the program is designed, about how the program's actions are supposed to achieve the intended outcomes. This 'logic model' is often not stated explicitly by those who run programs, rather it is assumed. The evaluator is required to draw out from the program staff how exactly the program is supposed to achieve its aims and assess whether this logic is holds true. In an HIV prevention program, for example, it might be assumed that educating people about HIV/AIDS transmission and risk, and safe sex practices will result in decreased infection rates. Research in South Africa, however, increasingly shows that in spite of increased education and knowledge, people still do not practice safe sex (Jacobs, 2003). The logic of a program that relies on education as a means to get people to use condoms may thus be faulty. Explicating this logic can also reveal unintended or unforeseen consequences of a program, both positive and negative. Developing a logic model can also build common understanding amongst program staff and stakeholders about what the program is actually supposed to do and how it is supposed to do it, which is often lacking.

Implementation Assessment

Implementation assessment or process evaluation looks beyond the theory of what the program is supposed to do and instead evaluates how the program is being implemented. This evaluation determines whether the components identified as critical to the success of the program are being executed. This evaluation determines whether target populations are being reached, people are receiving the intended services, staff are adequately qualified, etc. Process evaluation should be on-going and measures used to evaluate whether the program is being implemented effectively should be regularly applied.

Impact/Effectiveness Assessment

Impact/effectiveness evaluation determines the causal effects of the program. This involves trying to measure if the program has achieved its intended outcomes. This can require using sophisticated statistical techniques in order to measure the effect of the program and identify the causal relationship between the program and the various outcomes. More information about impact evaluation can be found in the section 'Determining Causation'.

Efficiency Assessment

Finally, cost-benefit or cost-effectiveness analysis assesses the efficiency of a program. Here, evaluators outline the benefits and costs of the program for comparison. While a program that is not efficient characteristically has a high cost-benefit ratio, an efficient program has a low cost-benefit ratio.

Determination of Causation

Explanation, an important end goal of research concerned with determination of the cause or causes of a given phenomenon (Kpolovie, 2010), is crucially important in utilitarian evaluation. For every phenomenon, there is a corresponding cause or causes. It is only through research that the factors, circumstances, antecedents and variables responsible for a given event or phenomenon can be identified. Explanation is given for an event when *why* and *how* an event occurred and *what* caused it are given. That is, a phenomenon can only be said to have been explained if causal inference has been established, a foremost concern of utilitarian evaluation research design. The conclusive or unquestionable establishment of causal inference or explanation demands three conditions (Kpolovie, 2010):

Evidence of co-variation of events

To correctly infer causal relationship between two variables or phenomena, both variables must co-vary (i.e., vary together). That is, changes in one of the variables or phenomena must directly result in changes to the other variable or phenomenon. Two variables can be said to co-vary only when an increase in the value of one of them is accompanied by a corresponding positive or negative change, or increase or decrease, in the value of the other. This occurs, for example, when an increase in the price of fuel results in transportation fare hikes. In a research setting, co-variation is determined by obtaining enough evidence that the manipulation of an event or variable creates proportional change in the other variable.

Co-variation can either be *deterministic* or *probabilistic*. *Deterministic co-variation* occurs when manipulation of only one variable produces observable effects of similar magnitude in another variable. In this case, the independent variable is said to be a necessary and sufficient condition for the observed effect in the dependent variable. *Probabilistic co-variation* occurs when the effect in the dependent variable is a function of the simultaneous manipulation of two or more independent variables. In this case, none of the independent variables can be deduced or inferred from the other and none of the independent variables is by itself a sufficient cause of the dependent variable, though all are collectively necessary for producing the effect on the dependent variable (Kpolovie, 2010). Multiple regression and two-way analyses of variance are the statistical tools that can help to understand this scenario. Co-variational evidence alone, however, is not enough reason for causal inference to be made.

Time-order relationship

The time-order relationship means that to correctly infer a causal relationship between independent and dependent variables, the independent variable must always occur before the dependent variable. That is, any change that occurs in the dependent variable when the independent variable(s) has/have not been operated on or manipulated cannot be attributed to the independent variable(s). The time-order relationship is anchored in the principle that the cause must occur before the presumed effect in every instance, trial or observation for causal inference to be made. If an aviation fare hike, for example, occurred before the federal government increased the price of fuel, it would be incorrect to claim that the increase in fares was caused by the increase in fuel price.

Elimination of confounding or extraneous variables

Causal inference can be made only when all confounding or extraneous variables have been eliminated in the course of manipulating the independent variable(s). Extraneous or confounding variables refer to the plausible alternative causes, other than the independent variables, of the effects observed on the dependent variable. Because confounding or extraneous variables are variables outside of the ones being studied, and are not only capable of, but actually cause changes in the dependent variable, they are termed *nuisance variables*. Nuisance variables must be eliminated or controlled in utilitarian evaluation research for the research to have *internal validity*. Findings of a study that did not eliminate nuisance variables adequately are of no use as the effects or changes in the dependent variables may have been caused by the extraneous, confounding or nuisance variables alone or jointly with the independent variables.

Perhaps, the greatest challenge in establishing causal relationship or causal inference is ensuring that the research has internal validity by eliminating or controlling all plausible extraneous, nuisance or confounding variables. Research without internal validity produces findings whose actual causes are not determinable. In such instances, the causes of the findings are erroneously or wrongly attributed to variables XYZ (the independent variables) when in fact, they were caused by variables ABC (the nuisance, confounding or extraneous variables). Effective randomization, matching, analysis of covariance, counterbalancing, control of subject effects, and control of experimenter effects are all excellent ways of eliminating extraneous variables and ensuring the internal validity of research (Kpolovie, 2010).

In summary, explanation of a phenomenon can only be achieved with provision of sufficient evidence of the co-variation of events, the establishment of a time-order relationship between changes in the independent and dependent variables, and elimination of all plausible extraneous, confounding or nuisance variables in determination of the occurrence of the phenomenon. Perhaps the most difficult part of evaluation is determining whether the program itself is causing the changes that are observed in the targeted population. Complex events or processes outside of the program may be the real cause of the observed outcome (or the real prevention of the anticipated outcome).

Causation is difficult to determine. One main reason for this is *self-selection* bias as people generally select themselves to participate in programs. In a job training program, for example, some people decide to participate and others do not. Those who do participate may differ from those who do not in important ways. They may be more determined to find a job or have better support resources. These characteristics may, in turn, be the actual cause of the observed outcome of increased employment, not the job training program.

If programs are able to use random assignment, they may find a strong cause-and-effect relationship. A program may, for example, randomly assign people to participate or not participate in the program, thereby eliminating self-selection bias. In this scenario, the group of people who participate would be the same as the group who did not participate. Inferential statistical tests of difference between groups, like the various types of t-test, ANOVA, and ANCOVA, are required for analyses of data in order to establish unquestionable causation (Kpolovie, 2011).

Since most programs cannot use random assignment, causation cannot be absolutely determined. Impact analysis can, however, still provide useful information. Outcomes of the program can be described, for example, using inferential statistical tests of association like regression, partial regression, and multiple regressions (Kpolovie, 2011). In this way, the evaluation can conclude that people who participated in the program were more likely to experience a given outcome than those who did not participate. If the program is fairly large, and there is enough data, statistical analysis can be used to make a reasonable case for the program by showing, for example, that other causes, while not completely implausible, are unlikely.

Requirements of a Good Sample

Inferential statistics entails a process of moving from the part to the whole, making observations of the *sample* and deductively drawing conclusions or generalizations from it to the entire population. In most cases it is not possible to observe the whole population and so knowledge of the whole comes from inferences or conclusions drawn validly from the study of a representative sample.

A sample is a portion of the research population that is meticulously drawn to adequately and proportionally represent all characteristics of the population, and from which data on the variables under investigation are collected and analysed for the making of valid inferences to the population. A portion drawn from the population must be truly representative of the parent population by proportionally accounting for all facets of the population before it can be considered and used as a sample. Only results from the observation of a representative sample can be rightly generalized to the population from which it was randomly drawn

While it is practically impossible in most circumstances for the evaluator to directly study every member of the population, it is generally not necessary to observe all the possible cases for a phenomenon to be understood. One does not have to eat all the fried peanuts in a basin in order to ascertain whether they were well and tastefully fried. Sampling allows for detailed observation of a portion of the population with regard to the phenomenon under study where the conclusions or results arrived can be mathematically generalized to the population with a satisfactory degree of certainty through suitable inferential statistical tests.

QUALITIES OF A GOOD SAMPLE

There are three qualities that a good sample must possess, namely, representativeness, accuracy and precision.

Representativeness

Emphatically, drawing of valid inferences from the sample to the parent population demands that the sample be obtained carefully and randomly for it to be a true **representative** cross section of individuals, events, elements, objects or characteristics in the population. McQueen and Knussen (2006) conclude that:

The key to grasping the significance of ‘whom do I study?’ problem lies in an understanding of the relationship between *samples* and *populations*, and an appreciation of how the phenomenon we are studying varies within the population... The hope is, of course, that what we learn about the sample will be true of the wider population, and the trick is to ensure that the individuals who comprise our sample are really representative of the population we wish to explore. However, irrespective of the size and scope of the population, the aim in studying a sample will always be representativeness: if our population contains both males and females, our sample must have a similar distribution; if there are varying ages in the population, our sample must reflect such age differences also. And if the proportion of males to females in the general population is uneven, then our sample too should ideally reflect this imbalance... [Worthy of note too is that] in many areas of cognitive psychology (e.g., testing visual acuity or response to audio signals, or memory) and in some medical research, it might be safe to assume that some measures will not vary as a result of, say, gender and personality, or regional location and culture (p. 91).

Accuracy

The second quality that a sample must possess is accuracy. A sample is said to possess accuracy when it is not biased positively or negatively in estimates of the values (parameters) or characteristics of the population from which it was randomly drawn. Positive bias occurs when sample elements overestimate the values of the population being investigated. Negative bias occurs when sample elements underestimate the characteristics of the population. Where a sample is large enough and is drawn randomly, possible overestimation and underestimation of population characteristics balance out or offset each other. An accurate sample is thus one that is large enough and randomly drawn to prevent overestimation and underestimation. Put differently, an accurate sample is one which is free totally

from bias and systematic variance. Systematic variance refers to variation in measurements due to known or unknown influences that cause scores to lean to one direction more than the other. Systematic variance is a form of an extraneous variable that must be prevented or eliminated.

Precision

The third quality of a good sample is precision. Precision requires that the numerical descriptors that describe the sample should not differ from those that describe the population. A sample is said to be precise or have precision when it is selected in such a way that best reduces sampling error (random fluctuation in the sampling process). Sampling error cannot be accounted for by systematic variance. Sampling error is composed of random fluctuations only and it is measured statistically with the standard error of estimate. The smaller the standard error of estimate is, the higher the precision of a sample. When the most appropriate sampling technique is used to obtain a sample, the standard error of estimate of the sample becomes smaller.

QUALITIES OF ACCEPTABLE MEASURING INSTRUMENTS

Instrument used in utilitarian evaluation and other evaluation studies must meet the requirements of a good measuring instrument. Instruments (for example, tests, questionnaires, and checklists) used in program evaluation must be as valid, reliable and sensitive as possible. A measuring instrument that is poorly chosen, or poorly conceived, or poorly developed can completely undermine the merit and worth of an evaluation, be it formative, process or impact assessment, by producing misleading estimates. It is only when outcome measures or instruments are valid, reliable and appropriately sensitive that the inputs, processes, or outcomes of a program can be credibly evaluated (Kpolovie, 2002; Kpolovie, 2010; Rossi, Lipsey & Freeman, 2004).

Validity of Measuring Instruments

The validity of a measurement instrument is the extent to which it measures that which it is intended to measure and nothing else. This concept itself can be difficult to accurately measure. In general use in evaluations, an instrument may be deemed valid if it is accepted as valid by stakeholders (funders, program administrators, etc.). Measuring instruments such as tests or questionnaires in utilitarian evaluation must be constructed in such a vigorous and rigorous manner that by all theoretical, abstract and empirical indications, they possess unquestionable validity.

A valid test on a specific attribute or social concept is that which perfectly, excellently or completely represents just that attribute or concept in direct congruence or concordance with the principles, assumptions, functions, paradigms, generalizations and theories of that social or psychological construct. To meet this requirement, each of the items in the test must be an accurate empirical indicator of the concept. Indeed, validity is the most important quality that a test must have. That is, without validity, a set of items, indicators or questions cannot be deemed a test. Scientific statements about the relationship among concepts or social science variables are distorted, obscured, and misleading if the concepts or variables are not validly measured. Indeed, what has plagued social science researchers most by preventing fuller description, explanation and prediction of phenomena is invalid or at least inadequate measurement of the phenomena or variables under investigation.

Almost all educational, psychological and social sciences variables or concepts are abstract, and are only made less abstract by their theories which postulate, propose or proffer certain complex patterns of interconnectivity among them. Consequently, for such constructs to be empirically measured with appreciably or satisfactorily high validity, each of the measurement instruments (tests) must be anchored firmly in the relevant theories on the concepts.

Thus, constructs like intelligence, personality, leadership style, motivation, conflict, aggression, and so on can only be validly measured when the measuring instruments are constructed in accordance with their corresponding theories and in line with the principles of testing. Instrumentation research is thus a rigorous, thorough and vigorous investigation aimed at the development of a valid test in which theoretical aspects of the concepts and measurement process are interweaved with the empirical or observable aspects of the concept and measurement process. Instrumentation research is used for the development and validation of a test by establishing unquestionable reliability and validity of the test for effective, efficient, accurate and feasible measurement of a particular variable, construct, trait, attribute, concept or phenomenon in the social or behavioural sciences. The measurement instruments in

utilitarian evaluation demand to be subjected to instrumentation research to ensure that they are individually of very high validity, reliability, and sensitivity. The three types of validity that must be established for a measuring instrument are content validity, criterion-related validity, and construct validity. Refer to Kpolovie (2010) for exceptional details.

Reliability of Measuring Instruments

For an instrument to validly measure a trait, it must as a preliminary step, be reliable in the measurement of that trait because while a reliable test may not possess validity, a valid test must possess reliability. The reliability of a measurement instrument is the extent to which the measure produces the same results when used repeatedly to measure the same thing. The more reliable a measure is, the greater its statistical power and the more credible its findings. If a measuring instrument is unreliable, it may dilute and obscure the real effects of a program, and the program will appear to be less effective than it actually is.

The aim of instrumentation research is the development of an instrument that accurately measures a particular social or psychological trait, construct, characteristic or behaviour of the examinees in the best manner without any unsystematic (random) or systematic fluctuation (error) in the instrument and characteristic under investigation. It is only when the instrument measures in this way that unquestionable quantitative descriptions of the respondents, individuals or examinees, in terms of the exact extent to which they possess and demonstrate the specific trait, can be made.

Of note, when the same measuring instrument is administered several times under the same circumstance to the same individual in the behavioural sciences, that individual usually obtains scores that randomly and unsystematically fluctuate to an extent (Kpolovie, 2002). This type of variation, due mainly to the measuring instrument rather than factors that are directly controllable by the examinee, denotes uncertainty in the quantitative description of the individual on the basis of the test. This unsystematic error or fluctuation in an individual's scores over several repeated testing means that in the behavioural sciences, one cannot completely depend on a single score obtained by an individual on an attribute that was measured once. To fully control or solve the problem of incomplete certainty in, and dependence on, a score obtained by an individual from the single administration of a test, instrumentation researchers must empirically establish the reliability of the measuring instrument. *Reliability* is a statistical coefficient that represents the extent or magnitude of consistency with which a measuring instrument or test faultlessly measures the trait it pertains to. Though psychological attributes are not generally measured consistently with 100 per cent accuracy, the exact degree of accuracy with which a test consistently measures the psychological characteristic that its items deal with, is the reliability coefficient of that particular test.

It is the responsibility of an investigator to develop a measuring instrument with very high reliability or consistency over time (over repeated tests), over two halves of the same test, over equivalent forms of the same test, and over the individual items that constitute the test for adequate and accurate measurement of a specific socio-psychological attribute. A set of items, questions or statements on an attribute is considered a test only when it empirically yields highly consistent or almost exactly the same scores for each individual internally (across the various items), repeatedly (across repeated testing over time), and externally (across parallel forms of the same test) on the trait. Thus, for a set of items to qualify as a test, it must be able to yield scores that are highly reliable so that accurate description and prediction of the examinees and other decisions can be made on its basis with certainty. Meaningful comparison of individuals with regard to a trait can only be made with a test that very reliably measures that trait.

Administering the same test to an individual two hundred times will yield several different scores. When plotted graphically, however, these scores will approximate a normal curve with about an equal number of scores below and above the mean. The mean score is the best estimate of the extent to which the individual possesses the attribute measured by the test. Deviation of the scores indicates that test scores are not completely reliable and that each score on a test is composed of an error component and a true component. A reliable test is one on which an individual obtains the same or very similar scores on a large number of separate administrations.

A similar normal curve indicating variation of test scores can be obtained when the same test is administered twice to one hundred individuals and the means of their pairs of scores are plotted on a graph. In this case, the test is said to be reliable if it places the group of examinees in the same rank order over the two administrations. Correlation of an individual's scores across the two occasions indicates the test's magnitude of consistency. When a large number of individuals having relatively stable differences with respect to an attribute write a test twice on that trait, the scores of some of the individuals differ such that the standard deviations of the two tests vary. Such variation is due to errors

of measurement. Decisions based on just one set of the test scores will be misleading except when it has been established that the test measures the trait in question in a highly reliable manner. There is therefore a great and urgent need for establishing the reliability of a test before correct and important decisions can be made on the basis of its scores.

Establishing the reliability of utilitarian evaluation measurement instruments must be done in strict accordance with *reliability theory*, which includes:

- Classical true-score-and-error-score model;
- Multifaceted analysis model of reliability;
- Conventional empirical methods of establishing reliability; and
- Domain sampling model.

Exhaustive explication of reliability theory can be found in Kpolovie (2010).

Sensitivity of Measuring Instruments

The principal purpose of the evaluation process is to measure whether the program under investigation has an effect on the social problem it seeks to redress; hence, the measurement instrument must be sensitive enough to discern these potential changes, where they indeed exist. A measurement instrument can be said to be insensitive if it contains items which measure outcomes that the program could not possibly affect, or if the instrument was originally developed for application to individuals (for example standardized psychological individual measures) rather than to a group setting (standardized psychological group tests). Ensuring sensitivity requires that a “group test” not be administered for individual testing, and an “individual test” not be administered for group testing. These factors may result in 'noise' which may obscure any effect that the program may have had.

It is only through the use of measures that adequately achieve the benchmarks of reliability, validity and sensitivity that credible evaluations can be produced. It is the duty of evaluators to produce credible evaluations, as their findings can have far reaching effects. A discreditable evaluation, for example, that is unable to show that a program is achieving its purpose when it is in fact creating positive significant change, may cause the program to lose its funding undeservedly and so must be guided against (Kpolovie, 2010).

CONCLUSION

Nigeria tends to frequently “reform” its educational system in a bid to fix the educational problems on which its national development hangs. This is based on the belief that no nation can develop beyond the quality of its education. Unfortunately however, these reforms (for example, the change from 6-5-2-3 to 6-5-4; 6-3-3-4; free universal primary education; housing of primary education under state/local government; 9-3-4; free and compulsory universal basic education; university education reforms; the National Open University; the drastic education reforms of 1999 to 2007; and the Education Reform Act 2007) have failed to convince the citizenry that our primary, secondary and tertiary institutions are performing as we would like them to.

These unsuccessful education reforms, not based on evaluation designs and evaluation models, led Kpolovie (2012) to describe Nigerian education reforms “as the perfect example of what education reform ought *not* to be”. It has become quite clear that the Nigerian education system is broken, crippled by a complex bureaucracy, flawed teacher policies, and misspent school funds, leaving the system in need of a sweeping reform that is drastically different from all previous reform attempts. Structural educational problems are now so deeply rooted in the country that more funding and small, incremental interventions are unlikely to make any difference.

As part of the need to overcome these challenges and fix Nigerian education, this article has looked at one facet of education reform and has concluded that such reform must be strictly based on suitable evaluation designs. This paper then examined 1) pseudo-evaluation designs; 2) questions/methods-oriented evaluation designs; 3) program improvement/accountability-oriented evaluation designs; and 4) social/advocacy-oriented evaluation designs. This paper recommended the use of final design, social/advocacy-oriented evaluation design. Within this design, this paper

has thoroughly examined, and advocated for the use of, a utilitarian evaluation design. A guide for executing education reform on the basis of utilitarian evaluation design was given.

Of all the designs for program evaluation, utilitarian evaluation design allows for the strictest quality assurance and control, greatest assessment of program merits and worth, highest levels of accountability and credibility, and the broadest continuous stakeholder participation in planning, implementation and assessment (Kpolovie, 2010; 2012). Utilitarian evaluation design best presents a rationale and framework for data-based information that enables program personnel to be responsible and accountable for their decisions and actions. Involvement of all stakeholders in the evaluation process ensures that their questions and needs are addressed and, in turn, engages them in the effective and efficient use of evaluation findings. Utilitarian evaluation design is also the most comprehensive in terms of attending to the context, inputs, processes, and outcomes of a program. It thus best meets the professional standards for program evaluation and for adhering to the democratic principles of a free society (Joint Committee on Standards for Educational Evaluation, 1994; 1981b). Utilitarian evaluation design is of the utmost value as it promotes positive psychology, positive individual traits, positive emotions and positive institutions. It is renowned for balancing the use of formative and summative as well as quantitative and qualitative methods in retroactively and proactively assessing a program's merits and worth for the greatest benefits of all stakeholders. The use of utilitarian evaluation design in the radical and revolutionary reformation of the Nigerian education system is indeed urgent in the revitalization, restructuring, refocusing, repositioning, and rectification of the nation's education. This is non-negotiable if Nigeria is to be developed as is widely desired.

REFERENCES

- Barbazette, J. (2006). *What is needs assessment?* Retrieved from January 1, 2012, from http://media.wiley.com/product_data/excerpt/57/07879752/0787975257.pdf
- Bennett, R. (2011). *How to design evaluation of educational and social programs*. Retrieved October 18, 2011, from http://www.ehow.com/how_7898422_design-evaluations-educational-social-programs.html
- Brown University (2011). *Rigorous evaluation*. Retrieved October 18, 2011, from http://www.lab.brown.edu/ae_rigeval.php
- Centres for Disease Control and Prevention (1999). *Framework for Program evaluation in Public Health*. MMWR 48(No. RR-11).
- Christie, C. A. & Fleischer, D. N. (2011). Insight into evaluation practices: A content analysis of designs and methods used in evaluation studies published in North American Evaluation-Focused Journals. *American Journal of Evaluation*. September 01, 2010. 31: 326 – 346. Retrieved October 19, 2011, from <http://aje.sagepub.com/content/31/3/326.abstract>
- Frisbie, R. D. (1986). *The use of microcomputer programs to improve the reliability and validity of content analysis in evaluation*. Doctoral dissertation, Western Michigan University: Western Michigan University Press.
- Greene, J. C., Caracelli, V. J. & Graham, W. F. (2011). *Toward a conceptual framework for mixed-method evaluation designs*. Retrieved October 19, 2011, from <http://epa.sagepub.com/content/11/3/255.short?rss=1&ssource=mfc>
- Haddad, W. D. & Demsky, T. (1995). *Educational policy-planning process: An applied framework*. Retrieved October 18, 2011, from http://www.unesco.org/education/pdf/11_200.pdf
- Harrall, A., Burt, M., Hatry, H., Ressman, S., & Sabol, W. (2011). *Evaluation strategies for human services programs*. Retrieved October 19, 2011, from http://www.ojp.usdoj.gov/BJA/evaluation/guide/documents/evaluation_strategies.html.
- HKU (2011). *Evaluation in social work practice: An overview*. Retrieved October 21, 2011, from <http://web.hku.hk/~hrnwlc/introsocwork/notesevaluat.htm>
- House, E. R. (2009). *Publications: IMA team, David Streatfield and publications*. Retrieved May 21, 2013, from <http://www.informat.org/publications/ernest-r-house.html>
- Jacobs, S. (2003). Child and Family Program Evaluation: Learning to Enjoy Complexity. *Applied Developmental Science*. 7(2), 62-75.
- James Bell Associates (2009). *Evaluation brief: Selecting an evaluation approach*. Retrieved January 1, 2012, from <http://www.jbassoc.com/reports/documents/evaluation%20brief%20-%20selecting%20an%20evaluation%20approach%20v5.pdf>

- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards. How to assess evaluations of educational programs*. Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (1981b). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.
- Knol (2012). *Terminology in evaluation: Approaches, types (purposes), methods, analysis techniques and designs*. Retrieved January 1, 2012, from <http://knol.google.com/k/terminology-in-evaluation-approaches-types-purposes-methods-analysis-techniques#>
- Kpolovie, P. J. (2002). *Test, measurement and evaluation in education*. Port Harcourt: Emhai Printing and Publishing Co.
- Kpolovie, P. J. (2010). *Advanced research methods*. Owerri: Springfield Publishers Ltd.
- Kpolovie, P. J. (2011). *Statistical techniques for advanced research*. New Owerri: Springfield Publishers Ltd.
- Kpolovie, P.J. (2012). *Education reforms without evaluation designs: Nigeria at risk*. Owerri: Springfield Publishers Ltd.
- Lichtenstein, R. (2011). *Quality evaluation designs: Education research, evaluation, policy*. Retrieved October 18, 2011, from <http://qualityevaluationdesigns.com/>
- Luca, P. (2010). *Types of evaluation design*. Retrieved October 21, 2011, from http://www.ehow.com/list_7617228_types-evaluation-design.html
- McNamara, C. (2011). *Basic guide to program evaluation (including outcomes evaluation)*. Retrieved January 1, 2012, from <http://managementhelp.org/evaluation/program-evaluation-guide.htm>
- McQueen, R. A. & Knussen, C. (2006). *Introduction to research methods and statistics in psychology*. Harlow, England: Prentice Hall.
- Miller, D. C. & Salkind, N. J. (2002) *Handbook of Research Design & Social Measurement*. Australia: SAGE.
- Murtala, B. (2011). *The case for Nigerian unity - Part 3: Education reform*. Retrieved October 18, 2011, from http://www.elombah.com/index.php?option=com_content&view=article&id=6472:the-case-for-nigerian-unity-part-3-education-reform&catid=79:barack-murtala&Itemid=37
- NOAA (2011). *Design evaluation for education projects*. Retrieved October 18, 2011, from <http://wateroutreach.uwex.edu/use/documents/NOAAEvalmanualFINAL.pdf>
- Ohio Department of Education (2011). *Ohio education reform*. Retrieved October 18, 2011, from <http://www.ode.state.oh.us/GD/Templates/Pages/ODE/ODEDetail.aspx?page=523>
- Porter, J. (2011). *Lesson plan for education reform*. Retrieved October 18, 2011, from http://www.businessweek.com/bschools/content/dec2006/bs20061214_250756.htm
- Potter, C. (2006). Program evaluation. In, M. Terre-Blanche, K. Durrheim & D. Painter (Eds.), *Research in practice: Applied methods for the social sciences*. Cape Town: UCT Press. Pp. 410-428.
- Rossi, P. H., Lipsey, W. M., & Freeman, H. E. (2004). *Evaluation: A systematic approach*. Thousand Oaks, CA: SAGE.
- Rouda, R. H., & Kusy, M. E. (1995). *Needs assessment: The first step*. Retrieved November 28, 2012, from http://alumnus.caltech.edu/~rouda/T2_NA.html
- Sanctuary Web Team (2011). *Design an evaluation Tool: Evaluation design checklist*. Retrieved October 18, 2011, from <http://sanctuaries.noaa.gov/education/evaluation/design.html>
- Siy, E. A. (2007). *Decision analysis as utilitarianism and standards setting as categorical imperatives: an industrial engineer's view on ethical theories in systems design*. Retrieved October 21, 2011, from <http://www.dlsu.edu.ph/research/journals/dlsuengi/pdf/vol1no2/02.pdf>
- Stanford Encyclopedia of Philosophy (2009). *The history of utilitarianism*. Retrieved October 20, 2011, from <http://plato.stanford.edu/entries/utilitarianism-history/>
- Stufflebeam, D. L.; Madaus, G. F. & Kellaghan, T. (Eds.), (2002). *Evaluation models: Viewpoints on educational and human services evaluation*. New York: Kluwer Academic Publishers.
- Sun Associates (1998). *Educational technology evaluation designs*. Retrieved October 19, 2011, from <http://www.sun-associates.com/eval/evalprop.html#anchor512131>
- The Design-Based Research Collective (2011). *Design-based research: An engineering paradigm for educational inquiry*. Retrieved October 18, 2011, from <http://www.designbasedresearch.org/reppubs/DBRC2003.pdf>

- The World Bank Group (2011). *Evaluation designs*. Retrieved January 12, 2011, from <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20188242~menuPK:415130~pagePK:148956~piPK:216618~theSitePK:384329,00.html>
- Trochim, W. M. K. (2006). *Introduction to evaluation*. Retrieved January 14, 2012, from <http://www.socialresearchmethods.net/kb/intreval.html>
- University of Pennsylvania (2007). *Positive Psychology Center*. Retrieved January 1, 2012, from <http://www.ppc.sas.upenn.edu/>
- University of Wisconsin-Madison (2011). *Culture, cognition and evaluation of STEM higher education reform: A mixed model*. Retrieved October 18, 2011, from http://www.wcer.wisc.edu/projects/projects.php?project_num=521
- Van-der-Riet, M. (2009). *The production of context: using activity theory to understand behaviour change in response to HIV and AIDS*. KwaZulu-Natal, Pietermaritzburg: University of KwaZulu-Natal Press.
- Wikimedia Foundation Inc. (10 November, 2011). *Program evaluation*. Retrieved January 1, 2012, from http://en.wikipedia.org/wiki/Program_evaluation
- Wikimedia Foundation Inc. (2011). *Educational evaluation*. Retrieved October 19, 2011, from http://en.wikipedia.org/wiki/Educational_evaluation
- Wikimedia Foundation Inc. (2011a). *Evaluation Approaches*. Retrieved October 20, 2011, from http://en.wikipedia.org/wiki/Evaluation_approaches
- Wikimedia Foundation Inc. (5 October, 2011). *Evaluation*. Retrieved October 20, 2011, from http://en.wikipedia.org/wiki/Evaluation_research

 © IJSRE

ⁱ Peter James Kpolovie is an Associate Professor in the Department of Educational Psychology, Guidance and Counselling, University of Port Harcourt, Nigeria. He earned his Bachelor of Education in 1991, his Masters of Education in 1996 and PhD in 2002 from the University of Port Harcourt. He specialises in education testing and measurement procedures. His PhD dissertation ‘Validation and standardisation of culturally fair intelligence tests for use in Nigeria’ was recognised with Nigeria University Commission best doctoral thesis award, the highest academic award in Nigeria. He is also the recipient of the Common Wealth Scholarship Award of Excellence. He has published extensively on intelligence testing, the application of information technology to data analysis, approaches to improved learning, and educational statistics and research methods.

ⁱⁱ Nwachukwu Prince Ololube, Ph.D., is a Senior Lecturer in the Department of Educational Foundations and Management, Faculty of Education, Ignatius Ajuru University of Education, Port Harcourt, Rivers State, Nigeria. He hold a PhD in Education and Teacher Education with focus in *Educational Management and Planning/Curriculum Studies* from the University of Helsinki, Finland. In addition, a postgraduate Diploma in Human Resources Management, Masters of Education in Educational Management and Planning, and a Bachelors of Science Education in Political Science. His research focuses on school business administration and management, curriculum studies and design, early childhood education, institutional management and leadership in higher education, education effectiveness, instructional effectiveness and quality improvement, ICT in education, adult and non-formal education, and research methodologies. Wherever I am and have always been, I have always displayed the same enthusiasm and dynamism toward inspiring and motivating my students. I challenge them to think critically and independently. I enjoy teaching as much as I do research. Dr. Ololube has published 6 textbooks, edited 3 books, presented at various international conferences, and contributed chapters to a number of books and encyclopedias. In all, Ololube has authored or co-authored more than seventy publications. Ololube’s professional contributions to the academic community include: Editor-in-Chief, International Journal of Scientific Research in Education (IJSRE); Editor, Online Journal of Education Research (OJER); Editorial Board Member, International Journal of Economics, Education and Development (IJEED); Review Board Member, International Editorial Review Board Member, International Journal of Management in Education (IJMIE); International Journal of Information and Communications Technology Education (IJICTE); Editorial Board Member Journal of Information Systems Education (JISE) and more. A selection of my publications and profile are available online at www.ololube.com