Evaluating the Verification Processes of Qualitative and Quantitative Methodology in Social/Behavioral Science: Does head-to-head comparison clarify the relationship or muddy the waters?

YASUDA, Tomoyuki

(出版者 / Publisher)
法政大学キャリアデザイン学会
(雑誌名 / Journal or Publication Title)
生涯学習とキャリアデザイン / 生涯学習とキャリアデザイン
(巻 / Volume)
15
(号 / Number)
1
(開始ページ / Start Page)
3
(終了ページ / End Page)
25
(発行年 / Year)
2017-11
(URL)
https://doi.org/10.15002/00014253

Evaluating the Verification Processes of Qualitative and Quantitative Methodology in Social/Behavioral Science:

Does head-to-head comparison clarify the relationship or muddy the waters?

Tomoyuki Yasuda
Associate Professor, Faculty of Lifelong Learning and Career Studies, Hosei University

Introduction

Social and behavioral science research based on the qualitative and quantitative methods has a long history of debates. A variety of arguments originated both from qualitative and quantitative traditions (e.g., Campbell, 1984; Cronbach, 1982; Guba & Lincoln, 1994; Reichardt & Cook, 1979). Fundamental to the debates were the differences in philosophical and methodological grounds underlying the qualitative and quantitative framework and whether or not different frameworks were compatible in application. The debates also existed as to what are the advantages and disadvantages associated with the two frameworks when being adapted to social/behavioral science research.

In the past, it was argued that the qualitative and quantitative research frameworks were grounded in two different paradigms and thus cannot be judged against the same standards (e.g., Guba, 1978; Patton, 1975). For instance, primary focus of quantitative research is often placed on seeking for truth, while meaning and understanding were valued the highest in qualitative research (e.g., Creswell, 1998). Standards for seeking truths are most likely to be different from those for meaning and understanding. Principles of incommensurability between paradigms (e.g., Kuhn, 1970), furthermore, seem to explain the incongruity between the qualitative and quantitative methods. Two methods under different paradigms can never be applied together because the paradigms are different to begin with. It is the paradigms that determine what is important, reasonable, and legitimate methods (Reichardt & Cook, 1979). Nevertheless, both qualitative and quantitative methods are more recently seen as useful tools for capturing different aspects of reality (e.g., program processes and outcomes), as addressed in the mixed-method research and evaluation designs. Given the increased applications of qualitative and quantitative methods,

separately or altogether, it is important that these two methods be compared from the theoretical and philosophical grounds and, more importantly, from the application grounds.

The main purpose of this paper is to compare and contrast the verification processes of qualitative and quantitative approaches in conducting social/behavioral science research. For that purpose, the present paper first discusses concisely about the characteristics of hierarchical relationships among paradigms, methods, and verification processes. The paper then examines principles and verification processes of the two mainstream methodology in the quantitative research traditions, namely the experimental and correlational approach. These approaches are dominant in social/behavioral science research. This paper then reviews types and characteristics of the validity that are discussed in qualitative research. Finally, it was examined how the verification process of qualitative research could be similar to, or different from, that of quantitative research. [1)]

Hierarchical Relationships among Paradigms, Methods, and Verification Processes

"Qualitative vs. quantitative debates" lies in the assumptions that methods are legitimately linked to the respective paradigms (Reichardt & Cook, 1979). Since paradigms are incommensurable or mutually exclusive by definition (e.g., Kuhn, 1970), methods based on one paradigm (e.g., quantitative methods) cannot be adapted for use in the other paradigm (e.g., qualitative methods). At more fundamental levels, the worldviews that the two paradigms subscribe are drastically different. For example, basic to the quantitative paradigm is the idea of staying objective, positivistic, deductive, outcome-oriented, and context-free, whereas the qualitative paradigm is usually thought of as being subjective, naturalistic, inductive, process-oriented, and dynamic (e.g., Reichardt & Cook, 1979). The difference can actually be expected from the origins of the two research paradigms. Qualitative research originated from the fields of anthropology and sociology and quantitative research from the natural and agricultural science (Richardt & Cook, 1979).

Grounded in these different philosophical outlooks, a variety of methodological frameworks were proposed by both qualitative and quantitative researchers. Most notably, experimental and quasi-experimental methods have been dominated in quantitative research (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002), whereas the naturalistic and contextual mode of inquiry have been playing a central role in qualitative research (e.g., Creswell, 1998; Denzin & Lincoln, 1994; Lincoln & Guba, 1985).

Differences in methodological frameworks between the quantitative and qualitative research provide with the diverse ways of looking at issues and problems. Similarly, the differences in the methodological applications suggest the unique or paradigm-specific ways of

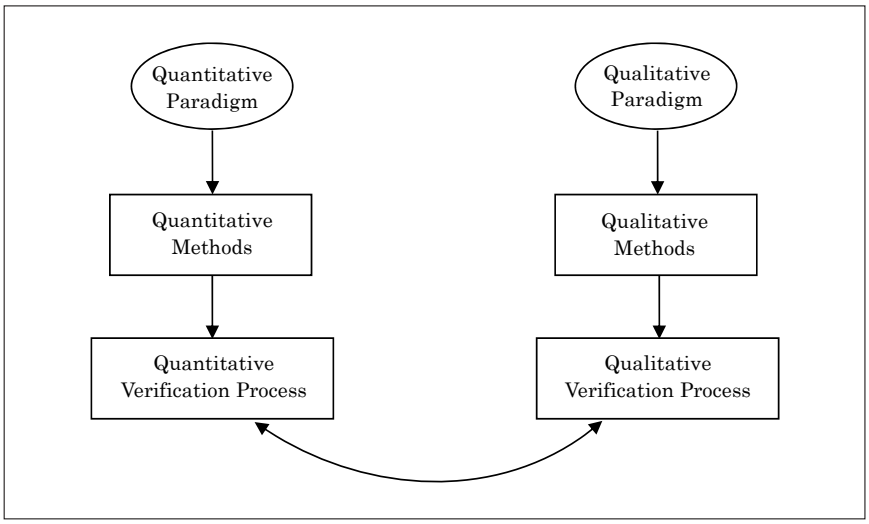

Figure 1. Quantitative and Qualitative Paradigms, Methods, and Verification Processes.

verification processes. Accordingly, Figure 1 shows the connectedness and hierarchical relationship of paradigms, methods, and verification processes.

Of central importance to this paper is the comparison of the verification processes between the qualitative and quantitative traditions. In general, verification procedures in quantitative research are well-known probably because of the introduction of the systems of internal and external validity along with reliability (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979). In contrast, verification procedures for qualitative research may not be documented well in the literature. This is partly due to the difficulties in defining validity as well as verification processes in qualitative terms. It is simply inappropriate, for example, to apply standards for verification processes used in the quantitative tradition (i.e., reliability and validity) to the qualitative traditions because ways to look at issues and problems, or paradigmatic characterizations more specifically, are different in the first place. In the sections that follow, verification processes in quantitative is reviewed.

Verification Processes in Quantitative Research

The quantitative framework includes numerous methods and techniques. Examples include randomized experiments, quasi-experiments, survey research, and multivariate statistical techniques. Accordingly, there are various types of validity as well as the standards for verification processes. Briefly, the traditional views of validity include the systems internal to the structure of a test (e.g., construct and content validity) and external to other variables such as concurrent and predictive validity (American Educational Research Association, American Psychological

Association, & National Council of Measurement in Education, 1999; see Nitko 2001 for details). Alternative aspects such as values, utility, and consequences are also discussed as of crucial importance to validity (e.g., see Messick, 1989 for detail).

Questions can be raised as to what are the verification processes that quantitative researchers engage. One relatively simple answer is that the researchers are to adhere to the methodological rigor. By methodological rigor, it was meant that the researchers follow adequate research procedures that are presumed to be able to "approximate the truth."

Approximating "Truth" in Experimental Research.

One important factor associated with the methodological rigor involves whether or not research is conducted under the framework of randomization or random assignment along with a control group. This approach, known as experimental tradition in quantitative psychology, was shown to maximize the evidence of internal validity. That is, the experimental research can provide the best answer to the question: "Did in fact the experimental treatments make a difference in this specific experimental instance?" (Campbell & Stanley, 1963, p. 5).

Internal validity is important in quantitative research because, with the high degree of it, the quantitative researchers can attain the observation of causal effect originated by the variable that is manipulated in the study. Accordingly, one of the most crucial aspects of the verification processes in experimental research is to design a research study where casual interpretations drawn from it are defensible. Thus, the exclusion of threats to the internal validity became "sin qua non" to the experimental researchers (Campbell & Stanley, 1963, p. 5).

Philosophy of falsification (e.g., Popper, 1959) plays a large role in the verification processes under the experimental design along with the quasi-experimental designs [2)] in the past (Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). Given that no theory or hypothesis can be confirmed under the systems of deductive logic (Popper, 1959), the next best approach researchers can take is to falsify or eliminate the competing hypotheses and explanations as much as possible (i.e., eliminate the threats to internal validity). In particular, randomization plays a significant role in eliminating some of the competing hypotheses (e.g., history, maturation, and selection) that are discussed by Campbell and Stanley (1963). Among all, the competing hypothesis that is always of crucial concern is the chance hypothesis (i.e, null hypothesis). In particular, statistical significant testing is usually adapted to rule out the chance hypothesis.

The bulk of descriptions in the work of Campbell and colleagues concerned the possible alternative explanations, as well as how to eliminate them, for the hope of "supporting" the researcher's hypothesis. Accordingly, the degree of methodological rigor can be equated with to follow a series of research designs (e.g., Pretest-Posttest

Control Group Design; Posttest-Only Control Group Design) that set forth by experimental psychologists (Campbell & Stanley, 1963; Cook & Campbell, 1979).

Another important verification process in the experimental research is to replicate the study. May simple it seems replications contain at least two theoretical advantages. First, direct evidence relevant to the purpose of study can be collected by replications, and hence contribute to the argument not only with respect to internal validity but also external validity.

Second and more important, replications do not involve any form of inductive logic. Reliance on deductive logic is consistent with the epistemology of the Popperian framework. Campbell and Stanley (1963) stated that "[W]hereas the problems of internal validity are solvable within the limits of the logic of probability statistics, the problems of external validity are not logically solvable in any neat, conclusive way," and hence, just like in David Hume' s truism, "induction or generalization is never fully justified logically" (p. 17). This is one of the very reasons why replications are seen as appropriate practice of science and are particularly preferred by some researchers as compared to the reliance on the statistical significant testing, which involves the use of inductive logic (e.g., Carver, 1978; Cohen, 1994).

In order to further advance the applications of the randomized experiments in social settings, Campbell (1988) proposed the notion of the experimenting society for social reform. Central to the experimenting society is that the causal effect observed in rather small settings (e.g., laboratories) can be examined in broader social settings as well. Campbell strongly believed that social reform can be conducted based on the framework of the experimental tradition. The construction of the notion of the experimenting society for social reform was originated mainly from his point of view in program evaluation.

In an ideal world of the experimenting society, Campbell believed (1988), the possible solutions (e.g., implementing a social or educational program) to the recurrent social issues and problems are experimented for the purpose of examining the "true" value of the solutions. A careful judgment on the study results can then be made. The new alternative solutions are tried out if the results have shown to be ineffective. In this spirit, Campbell valued the use experimental studies along with replications to social reform that can maximize the observation of the causal effect. When the experimental designs are usually implemented in evaluating any types of social and educational programs, the verification processes, implicit or explicitly, are operated under the "trial and error" of experiments. The principles of replications are also applied in the verification processes, so that researchers can determine if it is true, or false, that the experiment works in the different settings and populations than the ones that they studied.

Approximating "Reality" in Correlational Research.

Another stream of theoretical and methodological framework that is

available in verifying the quantitative researchers' work is correlational research. Presented as one of the two disciplines of scientific psychology (Cronbach, 1957) [3], the perspective that correlation researchers take differs substantially from that of experimental research. While experimental researchers are interested only in the variability that the researchers themselves create, correlational research examines the already existing variability between individuals, groups, and social settings (Cronbach, 1957). However, the predominant interests in rigorous tests of hypotheses and confident statement about causation using experimental research in psychology about half a decade ago prevented researchers from engaging in correlational research. As Cronbach (1957) explained:

> Everyone knows what experimental psychology is and who the experimental psychologists are. Correlational psychology, though fully as old as experimentation, was slower to mature. It qualifies equally as a discipline, however, because it asks a distinctive type of question and has technical methods of examining whether the question has been properly put and the data properly interpreted (p. 671).

There is an important reason as to why the experimental and correlational research are thought of as the two opposing end of continuum. Fundamental to the interests of experimental research is to control situational variables (which are the manifestation of individual difference) under tight control (e.g., randomization). In other words, individual variation or difference is treated as rather troublesome source of information and "is cast into that outer darkness known as 'error variance'" (Cronbach, 1957, p. 674). Accordingly, individual differences are the variability that should be reduced by any means in the experimental research tradition. In contrast, correlational researchers are "in love with just those variables, *error variances*, that the experimenter left home to forget" (Cronbach, 1957, p. 654, Italic added). Accordingly, perspectives that experimental and correlational researchers take with respect to error variance is totally different.

Given the interests in developing a model that can explain as much source of individual variations as possible, verification process based on the correlational research emphasize whether or not the proposed model adequately characterizes the complexity in real situations. Establishing causality guided by the limited scope in experimental tradition is just a part of the story in seeing the real world; what is needed is the multivariate conception of the world that can approximate the ontological complexity as close as possible.

Cronbach (1982) explained that the causal statements, hence focusing only on internal validity, are extremely restricted because causal conclusions are conditional, limiting the applications and use of the conclusions that were drawn from a study. Human behavior cannot or should not be reduced to a series of parsimonious causal laws, because "[W]hen he [or she] says that such-and-such a relation is true, 'other

things being equal,' [she or] he is speaking from the experience of having made a lot of things equal" (Cronbach, 1975, p. 121). As such, Cronbach's criticism mainly concerns the limited scope of internal validity, although this limited scope is sin qua non for the experimental tradition (Campbell & Stanley, 1963; Cook & Campbell, 1979).

There are several important conceptual frameworks that Cronbach and colleagues set forth that can guide the verification processes of correlational researchers. The analytical framework known as nomological network (Cronbach & Meehl, 1955) [4] has been, implicitly or explicitly, playing a central role in research involving the test of social, educational, and psychological theories under the framework of correlational research. Defined as the interlocking system of laws and concepts that is logically connected to a theory, the nomological network was developed in order to further advance the conceptualization of construct validity in the existence of ontological complexity in real world. Fundamental to the nomological network are the constructs and the links that connect these constructs. Cronbach and Meehl (1955) believed that to know what something was to form the lawful network of relationship within which it occurs. Thus, at least a necessary condition of a construct can be achieved if it is observed within a nomological network. In fact, central to the interest of quantitative researchers is to verify the newly-proposed theoretical framework with the preexisting framework that has already been developed. In the processes of verification, elaboration occurs if researchers learn more about the network. On the other hand, modification is sometimes needed by accepting the fact that there may be alternative ways of organizing the net which for the time being are equally defensible (Cronbach & Meehl, 1955).

Nevertheless, Meehl (1967) later on argued against the use of the nomological nework (that he himself proposed) in testing theories because of the stochastic characteristics associated with the testing of links of the constructs in the network. More specifically, he highlighted the distinction between the substantive theory underlying the networks and the statistical hypothesis from which a theory is derived. It is the test of statistical hypothesis using null hypothesis significance testing that is used for judging the quality of substantive theory. However, the law-like relationships that form the nomological network are "correlations, tendencies, statistical clusterings, increments of probabilities, and altered stochastic dispositions" and thus the nomological network in social science is "at best an extension of meaning and at worst a misleading corruption of logician' s terminology" (p. 813). More critically, he argued against the use of null hypothesis testing in social science and education, in which Carver (1979) and Cohen (1994) would probably concur, as follows:

> "Sir Ronald [Fisher] has befuddled us, mesmerized us, and led us down the primrose path. I believe that the almost universal reliance on merely refuting the null hypothesis as the standard method for corroborating substantive theories in the soft areas

> is a terrible mis-take, is basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology" (Meehl, 1978, p. 817).

Although these drawbacks associated with the nomological network were acknowledged about forty years ago, the idea of nomological network, implicitly or explicitly, has been serving as the backbone of the applied quantitative research since its inception.

Another framework that has been of central importance to the work of correlational researchers is the idea of maximizing external validity or generalizability. Although the importance of internal validity and causal explanations has usually been viewed as acknowledged among experimental and correlational researchers alike, Cronbach and colleagues (e.g., Cronbach, 1982; Cronbach, Ambron, Dornbusch, Hess, Hornik, Phillips, Waker, & Weiner, 1980) emphasized the central roles that external validity play in quantitative research in general and program evaluation in particular. For the purpose of shifting attentions away from the simple ways of making things equal by randomization, Cronbach specified the analytic framework that systematizes complex elements that need to be taken into consideration in program evaluation (e.g., unit of analysis, treatment types, and types of population and settings) [5] .

The specification of this conceptual framework most notably in program evaluation, as well as some measurement framework known as Generalizability theory (e.g., Cronbach, Gleser, Nanda, & Rajaratnam, 1972), basically points to the importance of specifying the sources or variables that can approximate the reality of ontological complexity as close as possible, so that the results are generalizable or have a high degree of external validity. The multivariate analytic framework, which was "[P]erhaps the most valuable trading goods the correlator can offer in return" (Cronbach, 1957, p. 676), can approximate the reality better by examining multiple sources of variations. As such, multivariate modeling became one of the most popular ways to verify the work done in correlational research.

Defining Validity in Qualitative Research Paradigm

Fundamental to the verification process of the quantitative research was to ensure the degree of methodological rigor (e.g., randomization) or the adequacy of multivariate conceptualization. In what way, then, the verification processes in qualitative inquiry can be conducted? Adherence to the methodological rigor, along with examining the assumptions of "your model," is considered important in qualitative research as well (e.g., Creswell, 1998). Nevertheless, methodological rigor is defined and operationalized differently in the qualitative research scheme.

Criteria against which quantitative research is judged are mostly positivistic in nature. The methodological rigor and adequate (statistical) modeling are thus of central importance to collect the evidence of reliability, validity, and generalizability

(e.g., Denzin & Lincoln, 2000). These are mainly for the purpose of maximizing the approximation of "truth" and "reality."

On the other hand, the main purpose of the methodological rigor in qualitative research is to maximize "meaning" and "understanding" (Mishler, 1990; Maxwell, 1996). As a result, the degree of credibility, correctedness, and trustworthiness, rather than reliability and validity, serve as basis for the verification processes in qualitative research (e.g., Creswell, 1998). Other factors associated with credibility involve the authenticity of a study (Lincoln & Guba, 1985). Emphasizing the importance of clarifying the concept of validity in qualitative research, for example, Maxwell (1996) discussed that:

> [T]he view that methods could guarantee validity was characteristic of early forms of positivism, which held that scientific knowledge could ultimately be reduced to a logical system that we securely grounded in irrefutable sense data. This position has been largely abandoned by philosophers, and methodologists are also becoming increasingly aware of the problems that this view creates. Validity is a goal rather than a product; it is never something that can be proven or taken for granted. Validity is also relative: it has to be assessed in relationship to purposes and circumstances of the research, rather than being a context independent property of methods or conclusions (Maxwell, 1992). Finally, validity threats are made implausible by *evidence*, not methods; methods are only a way of getting evidence that can help you rule out these threats (p. 86).

The possible differences about the definition and operationalization of validity can easily be expected from the incommensurability between qualitative and quantitative paradigms. As Figure 1 showed, it is the paradigm that determines the appropriate methods to be utilized; it is the method that becomes a means to obtain the evidence, which is then evaluated against the respective criteria of validity.

Credibility, Transferability, and Extrapolation

For the purpose of further clarifying the definition of validity in qualitative research, Lincoln and Guba (1986) redefined the validity in qualitative research in comparison with quantitative research. For example, internal validity in quantitative framework was defined as "credibility"; external validity was defined as "transferability."

First, credibility, or degree of credibility more accurately, in qualitative research is obtained by the prolong engagement and persistent observation in the field. Elements of credibility also include the integration of multiple perspectives by triangulations, peer debriefing or solicitations of feedback about the collected data and interpretation from the people that are studied, or participants (Guba & Lincoln, 1989).

On the other hand, transferability is defined as the degree to which the study findings can be transferred to other settings

and requires rich descriptions of data (Guba & Lincoln, 1989). Because words serve as the data in qualitative research (as numbers serve as data in quantitative research), the richer or thicker the description the better. One of the main purposes for engaging thick description is to place the participants in context as much as possible. Contexts obviously are changeable and variable on the regular or irregular grounds, and thus the examination of transferability are essential in generalizing the study findings to the contexts or populations that may be different in characteristics.

Interestingly, similar yet more specific definition of transferability was offered by Cronbach (1982) who concerns human behavior embedded in contexts. In describing external validity, Cronbach (1982) employed the concept of extrapolation, rather than generalizability, for the purpose of reflecting a deliberate projection to a situation outside of the range where information was gathered. His point of view can be well-articulated via the use of his symbol systems (i.e., utoS, UTOS, and *UTOS), which lead to the totally different or even paradoxical definition of internal and external validity.

Briefly stated, internal validity (in the Campbell's sense) essentially concern whether or not a treatment works in a specific experiment and thus examine the link between "t" or (actual) treatment and "o" or (actual) observation (i.e., from "t" to "o"). This is in sharp contrast with the Cronbach's formulation of internal validity, which concerns the statement about UTOS made on the basis on utoS because "[T]he inference is internal to UTOS in the sense that utoS is a subset of observations from that universe" (Cronbach, 1982, p. 106). Accordingly, he viewed internal validity as "reproducibility" (i.e., from "utoS" to "UTOS"), but strangely enough this is external validity to the Campbell framework (Shadish et al., 1991).

Cronbach (1982) furthermore argued that, since the real purpose of generalization is to extrapolate or project to a situation outside of the range where information was gathered, quantifications via statistical summaries on the sample (i.e., utoS), which serves the estimates of UTOS, are usually not an adequate base for the inference about *UTOS. In other words, for Cronbach the extrapolation from UTOS and *UTOS is external validity. This is the type of validity that does not exist in the Campbell's formulation. Most notably, just like the qualitative researchers point to the need of transferability, Cronbach (1982) claimed the importance of qualitative knowledge in maximizing external validity, such that history, culture, folktale, should be incorporated in the examination of the difference between the studied sample and settings and the populations. Adequate degree of transformation or extrapolation can thus be made. Finally, he noted that the collecting evidence of external validity (i.e., extrapolation) is less systematic and stylized than statistical inference, and the range of extrapolation needs to be shortened in order to improve the credibility of extrapolation.

Integrating Reliability in the Validity Framework

Although reliability is often discussed separately from validity in quantitative research[6] and usually examined as such in applied research (e.g., examination of internal consistency reliability estimates based on Cronbach's α[7] or construct validity via factor analysis), reliability and related concepts are integrated holistically as trustworthiness under the qualitative verification processes (Lincoln & Guba, 1986). Thus, the concepts of reliability and objectivity discussed in the quantitative traditions are framed differently using the different terms and definitions. In particular, reliability and validity were defined, respectively, "dependability" and "confirmability" (Lincoln & Guba, 1986). Dependability and confirmability are thought of as the replacement of reliability and objectivity in quantitative counterpart because both dependability and confirmability are established throughout an auditing and monitoring of the research process (Creswell, 1998).

Rather new type of criteria, public disclosure, was also discussed as an important element of reliability and validity in qualitative research (Anfara, Brown, & Mongione, 2002). More specifically, it was argued that the traditional categories of reliability and validity failed to address the criteria against which the values of qualitative research are evaluated, namely, the privatization of the data and analyses (cf., public database in quantitative research).

Other Types of Validity

Additional types of validity were proposed by Maxwell (1996), namely, descriptive, interpretive, and theoretical validity. First, descriptive validity involves the assessment of the study evidence on the factual bases. The main purpose of examining this validity is to ensure accuracy and completeness of the data. Data collections (e.g., interviews and observations) based on the audio and video recordings are at least necessary for obtaining the accuracy and completeness at the data collection stages.

Second, interpretive validity concerns the appropriateness of interpretation during and after the data collection. As Rossman and Rallis (2002) describes, "[D] ata are filtered through the researcher' s unique way of seeing the world – his [or her] lens or worldview" (p. 36). Accordingly, imposing a researcher's own framework can hinder the accurate understanding of the perspective of the participants studied, such as the meanings that they attach to their words (Maxwell, 1996). Finally, all data interpretation can be bound by theoretical frame of reference. Third type of validity, theoretical validity, attempts to examine the evidence of theoretical justifications of data that are collected by paying particular attentions to discrepancy between theory and data.

"Threats to Validity" in Qualitative Research

Just like there are threats to internal/ external validity that can jeopardize the conclusions drawn in quantitative research,

there are also validity threats in qualitative research. Although validity threats are often treated as generic factors to be controlled in the experimental tradition, the threats for qualitative research are rather treated uniquely depending on the circumstances and contexts in which the study is embedded (Maxwell, 1996). The two general threats inherent to the application of qualitative research exist: Researcher bias and reactivity (Maxwell, 1996). Researcher bias can occur, in theory, at any stage of the investigation; yet it most likely occurs at the interpretation stage (e.g., interpretive validity).

Essential to cope with the researcher bias is to disclose it rather than eliminate it. As Maxwell (1996) argued, it is obviously impractical to deal with the problems of researcher bias by eliminating the researcher's background, theoretical frame of reference, preconceptions, and values. Standardization can be a possible solution especially for maximizing reliability; yet maximizing reliability may not be an appropriate solution in qualitative research. Maxwell (1996) noted:

> Qualitative research is not primary concerned with eliminating variance between researchers in the values and expectations they bring to the study, but with understanding how a particular researcher's values influence the conduct and conclusions of the study. Explaining your possible biases and how you will deal with these is a key task of your research proposal (p. 91).

Reactivity, or "the influence of the researcher on the setting or individuals studied" (Maxwell, 1996, p. 91), also becomes a threat to validity in qualitative research. The same principle applied here as in dealing with the researcher bias. Namely, instead of trying to eliminate researcher reactivity, the influences of it onto the research participants as well as researcher's interpretation are to be understood and used productively (Maxwell, 1996). Degree of relevance differs depending on the type of qualitative investigation. For example, reactivity is not commonly a problem for observations, whereas it is for interviews. Qualitative research in general and interviews and related method (e.g., focus groups) in particular are essentially interactive, and the researcher involves with participants face to face (Rossman & Rallis, 2003). In other words, the comments of informant are always a function of the interviewer and the interview situation (Maxwell, 1996). Leading questions by an interviewer, for example, are likely to be malpractice (e.g., Maxwell, 1996; Mishler, 1986). Judgmental terms such as *should* and *ought* in the interviewers' sentences also bias the data collection (e.g., Wolcott, 1990a).

Verifying Qualitative and Quantitative Research Findings

Validity issues including the threats to validity have been shown important in both quantitative and qualitative research. Grounded in distinct worldviews (e.g., "meaning and understanding" vs. "truth and reality"), moreover, qualitative and quantitative research define validity and

verification processes differently (e.g., credibility vs. internal validity). This final section discusses how verification processes inherent to the qualitative research tradition are similar to, or different from, those of the quantitative tradition.

Research under the quantitative framework not only values qualitative knowledge, but also values it highly. As discussed earlier, importance of qualitative knowledge was emphasized by Cronbach (1982) who claimed that the knowledge, such as history, culture, and folktale, play a large role in maximizing extrapolation. Furthermore, Campbell (1984) acknowledged that qualitative knowing comes always earlier than quantification such as in generating research hypotheses. For example, factor analysis, which is often regarded as one of the most qualitatively-oriented methods in quantitative research, is used for the purpose of developing measures or gaining some insights about hypotheses, which are then tested quantitatively. Acknowledging the utility of practical knowledge, furthermore, Campbell only believed in science because it is marginally more effective. It is effective because there are norms and systems about identifying and publicly adjusting threats to knowledge claim (Shadish, Cook, & Leviton, 1991).

What are the standards for verification process in qualitative research? In other words, how credibility of qualitative research can be evaluated in systematically and practically effective manners? Given the viewpoints of qualitative researches, what are the processes involving in the verification of qualitative findings? As Creswell (1998) admitted, it is not enough to gain perspectives and terms of validity in qualitative research; but rather these ideas should be translated into practice.

Patton (2002) formulated the systematic approach to enhance the conditions of qualitative analysis. Specifically, he recognized the controversy associating with the nature of analyses. He noted, "[S]tatistical analysis follows formulas and rules, while, at the core, qualitative analysis depends on the insights and conceptual capacities of the analyst" (Patton, 2002, p. 553). Accordingly, what is needed in qualitative analysis are the systems that can provide researchers with methodological rigor. Thus, high degree of credibility on the researchers' insights and flexibility is achievable.

"Hypothesis Testing" in Qualitative Research

Excluding the rival conclusions is one approach to strengthen the credibility of qualitative findings (Patton, 2002). Central to this approach is to generate possible alternative explanations against which the findings are tested. In particular, the inductive and deductive approaches can be taken for the purpose of generating the alternative explanations. Possible predispositions and bias of the researchers are thus examined systematically. More specifically, this approach focuses particular attentions on the two perspectives of rival hypotheses: a) identification of other ways of organizing and interpreting the data that might lead to different results and conclusions (i.e., inductive); and b)

identification of other possibilities and examination of those possibilities that can be supported by data (i.e., deductive) (Patton, 2002).

The latter approach also resembles the verification under the framework of negative case analyses (e.g., Creswell, 1998; Maxwell, 1996; Patton, 2002). Essential to the negative case analysis is for the qualitative researchers to develop a working hypothesis in light of negative or discrepant results (Creswell, 1998). Optimal practice in the negative case analysis is the interpretations that fit the data along with the discrepant results, so that readers can evaluate and draw their own conclusions (e.g., Wolcott, 1990b).

The examination of alternative conclusions in qualitative research is indeed quite similar to the formulation and exclusion of the rival hypotheses or alternative models in quantitative approach. In particular, the experimental research attempts to exclude the rival hypotheses (e.g., history, maturation, and instrumentations) [8] via randomization (e.g., Campbell & Stanley, 1963). Moreover, correlational research often attempts to develop the alternative model(s) that may better explain the data at hand.

What may be different, however, is that the qualitative research engages in testing alternative conclusions less systematically than the quantitative research does. For example, the experimental as well as quasi-experimental researchers usually deal with the rival hypotheses that are well-defined in advance (e.g., threats to internal validity). Moreover, they do so by following the established procedures discussed in the literature such as using randomization and/or control group (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979). On the other hand, given the interests of qualitative research are on the meaning and understanding of the real world (cf., causation and correlation in quantitative research), the alternative conclusions can be highly contextual and situational in nature. Perhaps more importantly, in quantitative research the rival hypotheses or alternative models are essentially known *a priori*, while the manners in which qualitative research attempts to exclude the alternative conclusions are *a posteriori*. In particular, Maxwell (1996) explained that the qualitative researchers:

> rarely have the benefit of formal comparisons, sampling strategies, or statistical manipulations that "control for" the effect of particular variables, and they must try to rule out most validity threats after the research has begun, using evidence collected during the research itself to make sense these alternative hypotheses implausible (p. 88).

Rather formal approach in excluding the rival explanations is based on the advocacy-adversary model (Wolf, 1975) or the modus operandi approach (Scriven, 1974). Applications of these approaches can be found mostly on the field of program evaluations (e.g., Patton, 2002). Central to these approaches, according to Patton (2002), are to form two separate groups, in which one group, named advocacy group, tries to find the information that is supportive of positive results (e.g., certain

program is effective), while another group called adversary group attempts to find the negative results such as the program is not effective and hence ought to be changed or terminated. By taking the judicial-like model in evaluating the quality of the findings, the approaches are regarded as criticism- and consensus-based approach to validity (e.g., Maxwell, 1992).

Methodological Triangulations

A commonly applied approach available for qualitative researchers to enhance the validity of qualitative findings is triangulation (e.g., Creswell, 1996; Maxwell, 1992; Patton, 2002). Defined as "collecting information from a diverse range of individuals and settings, using a variety of methods" (Maxwell, 1992), the triangulation process involves the use of multiple information sources. The main purpose of triangulation is to provide the corroborating evidence, and as such the theme and perspective from different angles can be emerged (Creswell, 1996).

In particular, four different aspects of triangulation were emphasized by Patton (2002): namely, the methods triangulation, resource triangulation, analyst triangulation, and theory triangulations. This formulation involves the use of multiple data collection methods (i.e., the methods triangulation). Within the same methods, different data sources can also be sought (i.e., the resource triangulation). Additional triangulations involve the multiple analysts for reviewing findings (i.e., the analyst triangulation) and the multiple theoretical perspectives for interpreting the data (i.e., the theory triangulation).

Central to a series of triangulation is to find the degree of convergence in the qualitative data of the same phenomena. First, the methods triangulation constitutes a form of comparative analysis with respect to data collections (Patton, 2002). Second, examination of the consistency obtained at different means and times within each qualitative method is also an important piece of information. Thus, the resource triangulation attempts to compare observations with interviews; comparing what informants say in public with what they say in private; comparing what they say about the same thing overtime; comparing interviews with the available documents and other written evidence (Patton, 2002). Third, employment of multiple analysts also helps to reduce potential bias and distorted interpretation associated with the single analyst. Accordingly, the main purpose of including the analyst triangulation in verification process is to check on bias in data collection and interpretation. Fourth, examining the same phenomena from multiple frames of references or theories can help uncover the people's words and behaviors that are inherently embedded in contexts and environment, hence the theory triangulation.

Quantitative researchers also engage in triangulation, but they do so differently from, and actually less often than, the qualitative researchers do. Indeed, one of the few related methods that was proposed in the field of quantitative psychology is the examination of convergent and

discriminant validity using Multi-Trait Multi-Method (MTMM) Matrix (Campbell & Fiske, 1959). Briefly, in the MTMM approach the construct of interests are examined via the multiple measures (i.e., multi-trait) and multiple methods (i.e., multi-method) of the same construct. Evidence of convergent validity can be obtained by finding the high correlations among the corresponding measures (e.g., three different measures of self-esteem) as well as corresponding methods (e.g., self-, teacher-, and mother-report) of the construct under investigation. On the other hand, evidence of discriminant validity can be obtained when there are low correlations among the non-corresponding measures and methods of the same construct. [9)]

The MTMM approach used in quantitative research is probably comparable to the combination of the methods/resource and theory triangulations in the qualitative counterpart. The methods/resource triangulation can provide integration of multiple sources of information and data collections such as observations and interviews. Furthermore, the theory triangulation can provide qualitative researchers with multiple ways of looking at the "construct" under investigation, that is, from "Multi-Trait" point of view. On the other hand, the analyst triangulation is most likely to be comparable to what quantitative researchers refer to reliability in general and internal consistency in particular. Quantifications are often done in the form of calculating the inter-rater reliability, which is used to measure the consistency of ratings among raters or judges, and this reliability indeed can be thought of close approximation of the analyst triangulation in qualitative research.

Credibility of "Instruments"

In quantitative research, measurement plays the most important role. Unless what researchers are measuring is "reliable," any conclusions drawn from a study cannot be trusted. Reliability of measurement device in quantitative research is crucially important, and this is most likely to be the reason why people sometimes say that reliability is a prerequisite to validity. Unreliability of measurement instrument, for example, can occasionally lead to the increased statistical regression in experimental research (e.g., Campbell & Stanley, 1963). Correlational research under the framework of the nomological network (e.g., Cronbach & Meehl, 1955) is meaningless unless constructs under investigation are measured reliably. For example, unreliability, or score unreliability more accurately, can introduce the attenuated correlations among constructs, and as a consequence the hypotheses developed under the network cannot be tested correctly both theoretically and statistically. This fact indeed points to the importance not only of score reliability but also construct validity in the nomological network (e.g., Cronbach & Meehl, 1955).

The investigator is the instrument in qualitative research, and thus the examination of "measurement error" in qualitative research is extremely difficult.

Any personal and professional information could affect data collection

(Patton, 2002). In order to ensure the validity of research findings in quantitative research, for example, relevant psychometric information, such as reliability coefficient (e.g., Cronbach's α) and validity evidence, are always presented in the quantitative research report; the information is usually reported under the section of instrumentation. In contrast, Patton (2002) argues that:

> a qualitative report should include some information about the research. What experience, training, and perspective does the researcher bring to the field? Who funded the study and under what arrangements with the researcher? How did the researcher gain access to the study site? What prior kno-wledge did the researcher bring to the research topic and study site? What personal connections does the researcher have to the people, program, or topic studied? (p. 566).

The in-depth information about the credibility of "instruments" or investigators can provide consumers of qualitative research with the perspectives under which measurement took place.

Patton (2002) further specified several problems of measurement processes in qualitative research. First, the anthropological literature discusses the problems of reactivity, and, for the purpose of minimizing them, the investigators need to engage in long-term observations. Prolong engagement that can be achieved by longitudinal observations helps investigators, and the people being studied have an opportunity to get used to each other in the study settings (Patton, 2002).

Second, bias and predispositions of the investigators can deteriorate the measurement in qualitative research. Researcher bias can be one of the crucial threats to the validity of qualitative research, and, as discussed earlier, attempts should be made on identifying and acknowledging the bias rather than eliminating it. Development of the section discussing validity issues in the selection of samples, data collection, personal bias that may increase the credible interpretations and conclusions. For example, researchers are often to develop the paragraph that could be entitled: How do I know what participants say is true and not just what I want to hear? (Maxwell, 1996). The researchers can thus discuss the issues of anonymity and confidentiality as well as power in personal and social relationships that may distort the conclusions. Such researcher's bias can actually be detected based on the triangulation results as to how she or he reaches to the conclusion (e.g., students and teacher reports and researcher observation). Similarly, issues as to "how an investigator might be wrong" could be addressed by discussing: How do I know what the participant says he or she does is true? (Maxwell, 1996). The field of quantitative research and educational measurement in particular has also witnessed the increased interests in dealing with the issues of test bias. Central purpose of the analyses of test bias is to identify and either exclude or modify the items containing the bias. Detection of bias in quantitative research is rather mechanical in that the statistical criteria in

judging the biases are usually applied.

Third, other undesirable factors, such as changes or incompetence of field investigators, can certainly affect the degree of validity in qualitative findings. Similarly, changing the instrument(s) during a study can produce uninterpretable results in quantitative research. Moreover, use of "sloppy" instruments in quantitative research produces meaningless results just as in the "garbage in garbage out" principle.

"Conceptual Factor Analysis" in Qualitative Research

Numbers are data to quantitative researchers. Quantitative researchers need to reduce the numbers into some related and theoretically meaningful clusters so that they can work effectively (e.g., reducing the number of parameters). This data reduction procedure is most commonly done via the procedure known as factor analysis. Since its invention (e.g., Spearman, 1904), [10] factor analysis has been a frequently-used statistical procedure in measurement and psychometrics to gauge the latent constructs (i.e., factors) under data. Factor analysis plays a fundamental role especially in correlational research because researchers first need to make sure the construct(s) under investigation can well be represented by the data at hand. More important is the theoretical justification of the construct(s). Both exploratory and confirmatory approaches (i.e., exploratory and confirmatory factor analysis) exist; a variety of statistical and conceptual criteria are applied in determining the number and characteristics of factors; and researchers rely both on their theoretical model and statistical criteria in applying factor analysis for research.

Words are data to qualitative researchers. How can qualitative researchers reduce the amount of data to make the data workable? Most often data reduction in qualitative research is done in a manner that may be described as "conceptual factor analysis" (cf., statistical factor analysis). The conceptual factor analysis is different from statistical one mainly because it is people, rather than computers, who engage in data reduction. Coding categories thus play a very important role for the conceptual factor analysis. As Bogdan and Biklen (1992) described, there are several steps in developing coding categories. Initially, data are inspected for regularities and patterns as well as for the topics the data cover. Researchers then write down words and phrases to represent these topics and patterns, which become coding categories. The coding categories are "a means of sorting the descriptive data you have collected (the signs under which you would pile the toys) so that the material bearing on a given topic can be physically separated from other data" (Bogdan & Biklen, 1992, p. 161).

Conceptual factor analysis is difficult mainly because, unlike the statistical counterpart, there can be lack of criteria against which the coding categories are determined. Development of the criteria, accordingly, becomes a crucial step in verification process in qualitative research. The qualitative data analysis and interpretation are defined respectively

as "the process of systematically searching and arranging the interview transcripts, fieldnotes, and other materials that you accumulate to enable you to come up with findings" and "developing ideas about your findings and relating them to the literature and to broader concerns and concepts" (Bogdan & Biklen, 1992, p. 147). In particular, Bogdan and Biklen described the importance of using the coding categories such as setting codes, situation codes, and activity codes. Coding categories based on differential perspective held by different informants are also discussed important. Indeed, the codes to qualitative researchers are the methods of factor extractions to psychometricians. Using the appropriate code and coding schema, qualitative researchers can interpret correctly what theme is emerged (cf. how factors are extracted) and how the theme are related or unrelated to other theme(s).

Reducing "Error Variance" in Qualitative Research

As discussed earlier, important to the experimental research is to minimize the error variance as much as possible. Correlational researchers, on the other hand, attempts to engage in multivariate conceptualization of the world so that the model can explain the data, or error variance more specifically, as much as possible. Indeed, the advanced multivariate procedures attempt to enhance the validity and verification process from data analytic perspectives by paying particular attentions on the error variance, which they are in love with (e.g., Cronbach, 1957). Examples include the application of popular multivariate techniques such as Structural Equation Modeling (SEM) and Hierchical Liniear Modeling (HLM). More specifically, measurement using SEM allows researchers to control error variance inherent to the measure of construct by separating the true variance from measurement error [11] that could attenuate the relationship between the construct and measurement. This was accomplished, in essence, by incorporating factor analysis in the regression model. Moreover, one of the crucial factors that lead to the development of HLM is to deal with the violation of independence assumption in error variance. HLM can effectively model the error variance in which the assumption of independence is violated because of the nested data structure. Accordingly, error variance plays the most important role in discussing the evidence of the validity in correlational research.

In what way "error variance" in qualitative research can be defined and reduced?

One of the cardinal principles responsible for enhancing the validity of qualitative research, according to Patton (2002), is to keep findings in context. In communicating findings, thus, it is very important that the purpose and limitations of the sample studied be clearly delineated, so that extrapolation of the findings is reasonable (Patton, 2002). On the other hand, failing to keep findings in contexts may produce the error variance in qualitative research. For the purpose of ensuring the methods and data in context, a variety of procedures can be applied. Member check, for example,

can be used to make sure the information as to if the data descriptions are consistent to what the informants said and meant. Adjustment can also be made based on the results of the member check. Of course, thick description also plays a central role in obtaining as much relevant information about context as possible. Patton (2002) argued that the design checks can serve to make sure appropriate methods and data are used to collect data. Given that context is situational, hence it is often changeable, furthermore, any sources of error originated from the flow of qualitative research (e.g., data collection, data reduction, and interpretation) should be examined in its appropriateness as design checks (Miles & Huberman, 1984). On the other hand, one may argue that definition of error and error variance inherent to quantitative investigation may not appropriately capture the characteristics of the error associated with qualitative research. Errors in quantitative research can be treated as static, while those in quantitative research may be dynamic and treated as such. The "liner" definition, where error variance is "left over" (i.e., residuals) from researchers' model, may also not be appropriate in qualitative research.

Concluding Remarks

The main objective of this paper was to examine the verification processes of qualitative and quantitative methodology in social and behavioral science. Accordingly, the paper looked at how validity is defined and examined both in quantitative and qualitative research. There are not only different points of view with respect to what validity is *across* the paradigms but there are also *within* paradigms (i.e., experimental and correlational research). Indeed, the philosophical underpinnings of correlational approach might be in close association with those of qualitative research more than those of experimental research. In particular, approximation of the ontological or ecological conditions inherent to human behavior was important in correlational research, so as the understanding and meaning of dynamics in contexts was essential to the qualitative tradition. Given the nature and aspects of validity that each approach is dealing with, verification processes also differed. While quantitative researchers follow a set of pre-existing standard to make their research rigorous, standards against which the validity of qualitative research is tested are not necessarily as clear-cut, primarily because qualitative researchers take rather dynamic point of view in verifying their work.

Endnotes

1) This paper is based on the response to one of the take-home questions "Compare processes used by qualitative and quantitative researchers to verify/validate their work" given by Dr. Mindy Kornhaber in the comprehensive exam of the educational psychology Ph.D. program at the Pennsylvania State University (May 20, 2004).
2) The designs that lack full experimental control (i.e., randomization and/or control group; Campbell & Stanley, 1963).
3) L. J. Cronbach later on incorporated the notion of aptitude treatment interaction (ATI) for the attempt of integrating, or go beyond the

dichotomization of, experimental and correlational approach.

4) Cronbach acknowledged the fact that most of the thoughts of this work come from Paul Meehl even though he was the second author of the article. Indeed, Cronbach later noted that the order of the authorship was determined by means of coin flip that Meehl insisted.

5) The symbolic characterization of this framework is known as "UTOS," representing unit, treatment, observation, and setting, respectively. Briefly, "U," or unit, represents population of persons about which a conclusion is sought; "T" or treatment denotes the plan for the program and its installation; "O" or observation represents the plan for collecting some type of data and measurement via quantification; and "S" or setting stands for situations in which the study is made. It is Cronbach's (1982) point that data collection usually takes place in the case specified by "utoS," in which lower-cases "uto" represents actual participants (u), actual treatment (t), and observation or data (o), whereas settings (S) remain upper-case since settings are unique to each study. Researchers want to know about the case specified as "*UTOS" (pronounced as Star UTOS), which represents a set of units not regarded fully represented of "U" (i.e., *U); a set of treatment that is different from actual treatment "T" (i.e., *T); observations or "O" that are not measured directly in the original study (i.e., *O); and the settings that are different from the particular study (i.e., *S).

6) There are some formulation of compatibility in reliability and validity, such as validity and reliability continuum. Also, the reliability and validity paradox exist in objective measurement.

7) More recent thoughts on Cronbach's α can be found in Cronbach (2004).

8) Other non-traditional types of alterative explanations (e.g., John Henry effect) exist (see Gall, Borg, & Gall for details).

9) Other approaches under different measurement framework examine issues of validity in quantitative research (e.g., M. Kane's sampling model of validity under Generalizability theory), although main focuses are, as the name suggests, generalization rather than triangulation. Moreover, although it may be a little bit different from how qualitative researchers use the triangulation methods, the comparisons of different methods and statistical procedures using the same data sources, simulated or real, are very frequently practiced in the quantitative field.

10) In 1901, K. Pearson also developed the methods of principal axes, focusing primary on the statistical rather than psychometric aspects of factor analysis.

11) More precisely, it is the unique variance that contains measurement error as well as systematic error.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Anfar, Jr. V. A., Brown, K. M., & Mangione, T. L. Qualitative analysis on stage: Making the research process more public. *Educational Researcher*, October, 28-38.

Bogdan, R. C., & Biklen, S. K. (1992). *Qualitative research for education*. Boston: Allyn & Bacon.

Campbell, D. T. (1984). Can we be scientific in applied social science? In R. F. Connor, D. G. Altman, & C. Jackson (Eds.), *Evaluation studies review annual* (Vol. 9, pp. 26-48). Beverly Hills, CA: Sage.

Campbell, D. T. (1988). *Methodology and epistemology for social science: Selected papers*. In E. S. Overman (Ed.).

Chicago: University of Chicago Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Carver, R. P. (1978). The case against statistical significant testing. *Harvard Educational Review*, 48, 378-399.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.

Cronbach, L. J. (1982). In praise of uncertainty. In P. H. Rossi (Ed.), *Standards for evaluation practice* (pp. 49-58). San Francisco: Jossey-Bass.

Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391-418.

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

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D.

F., & Weiner, S. S. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.

Denzin, N. K., & Lincoln, Y. S. (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational Research: An introduction*. White Plains, NY: Longman.

Guba, E. G. (1978). Toward a methodology of naturalistic inquiry in educational evaluation. *CSE Monograph Series in Evaluation No. 8*. Los Angeles: Center for the Study of Evaluation.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

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.

Kuhn, T. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

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

Lincoln, Y. S., & Guba, E. G. (1986). "But Is It Rigorous? Trustworthiness and Authenticity in Naturalistic Evaluation." In D. Williams (Ed.). *New Directions for Program Evaluation*, 30, 73-84. San Francisco: Jossey-Bass.

Maxwell, J. A. (1996). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.

Meehl, P. E. (1967). Theory-testing in psychological and physics: A methodological paradox. *Philosophy of Science*, 34, 103-115.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Assessment* (3rd ed., pp.13-103). Englewood Cliffs, NJ: Prentice Hall.

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis:* A sourcebook of new methods. Beverly Hills, CA: Sage.

Mishler, E. G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review*, 60, 415-442.

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

Nitko, A. J. (2001). *Educational Assessment of Students* (3rd Ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Patton, M. Q. (1975). *Alternative Evaluation Research Paradigms*. North Dakota Study Group on Evaluation monograph series. Grand Forks: Center for Teaching and Learning, University of North Dakota.

Popper, K. R. (1959). The Logic of Scientific Discovery. Routledge: New York.

Reichardt, C. S., & Cook, T. D. (1979). Beyond qualitative versus quantitative methods. In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 7-32). Beverly Hills, CA: Sage.

Scriven, M. (1974). Maximizing the power of causal investigations: The modus operandi method. In W. J. Popham (Ed.), *Evaluation in Education – Current applications* (pp. 64-84). Berkeley, CA: McCutchan.

Shadish, W. R., Cook, T. D., Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.

Shadish, W. R., Cook, T. D., Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin Company.

Spearman, C. (1904). General intelligence, objectively determined and measured. American Journal of Psychology, 15, 201-293.

Wolcott, H. F. (1990a). On seeking – and rejecting – validity in qualitative research. In E. W. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 121-152). New York: Columbia University, Teachers College Press.

Wolcott, H. F. (1990b). *Writing up qualitative research*. Newbury Park, CA: Sage.

Wolf, R. L. (1975). Trial by jury: A new evaluation method. *Phi Delta Kappan*, November.