

RELIABILITY AND VALIDITY: MISNOMERS FOR QUALITATIVE RESEARCH

Olive Yonge and Len Stewin

Validity and reliability, as indices of measurement, have been well developed in and for quantitative methods. Validity has been classified and defined in the following ways: face, content, predictive, concurrent, construct, internal and external. Reliability has been classified and defined as quixotic, diachronic, synchronic, external and internal (Duffy, 1985; Kirk & Miller, 1986; Knapp, 1985; Le Compte & Goetz, 1982). Essentially the terms validity and reliability refer to accuracy, consistency and equivalence in research that is designed for quantification in the natural sciences. Many tests have been and continue to be developed to prove validity and reliability. Not all quantitative researchers agree however, that the terms are used accurately. For example, Knapp (1985) in his article "Validity, reliability and neither" notes errors that are commonly made: describing internal consistency as a distinct and different property, when in fact it is an aspect of reliability; or, the citing of a correlation between parallel forms as evidence for validity. Knapp illustrates how some researchers incorrectly use these measures and he raises a second issue - that knowledge about reliability and validity is continuing to evolve.

The thesis of this paper is that the terms validity and reliability should not be applied to qualitative research methods. When Le Compte and Goetz (1982) wrote their, seminal article on how qualitative research has adhered to the canons of reliability and validity, what was their intent? They acknowledged that reliability is difficult to measure in qualitative research because of the nature of the narrative data and the involvement of the researcher in a change process; yet, they proceeded to force qualitative methods to fit criteria for external and internal reliability. Other qualitative researchers such as Goodwin and Goodwin (1984), Duffy (1985), Swanson-Laufman (1986), and Atwood and Hinds (1986) have followed suit, describing how their research addresses these measures.

Olive Yonge, R.N., M.Ed. is Associate Professor in the Faculty of Nursing at the University of Alberta. Len Stewin, Ph.D. is Professor in the Faculty of Educational Psychology at the University of Alberta, in Edmonton.

There are different historical origins for quantitative and qualitative enquiry, different interpretations of the words "meaning" and "truth" and different emphases on methods of data collection. As such, it is difficult to understand why terms associated with measurement theory would be used to describe research that is embedded in rather than removed from context and is a "systematic study of the world of every-day experience" (Swanson & Chenitz, 1982, p.243). For example, because the data analysis and collection stage proceed simultaneously when using grounded theory, at what point should the terms be applied? As well, how could a panel of experts validate data when they were not part of the social situation that generated them (Stern, Allen & Moxley, 1984)? One could speculate that the terms have been applied for the following reasons: thoughtlessly; to convince quantitative researchers about the rigour of qualitative methods; to fill a void, as a way of describing credibility; or, thoughtfully, because no essential differences between the two methods are identified. It is a belief of the authors that there is an essential difference between the two methods of enquiry, and that the question of rigour in qualitative methods can be addressed in ways other than the traditional validity and reliability measures referred to in quantitative research.

Essential Differences

Research is a personal and social process (Allender, 1986). Over time, this subjective process mirrors and responds to political and social pressures. Reality no longer is static but is dependent on the context and perceptions of the actor, each having his own reality. This view of reality is part of a new paradigm that conflicts with the established positivistic, reductionist, deterministic paradigm that is concerned with quantification. This new paradigm acknowledges that, as researchers, "we interact with and are an integral part of what we study" (Allender, p. 185); this places a tremendous onus on the researcher to be responsible and ethical.

Smith and Heshusius (1986), using different descriptors, outline ways in which qualitative approaches arose as a reaction against positivism. Citing the observations of Dilthey that there is a fundamental difference in subject matter between the natural and social sciences, they observe that other researchers have ignored this division. In fact they accuse certain qualitative researchers of using a "variation in technique" in a quantitative framework. Miles and Huberman (1984), criticized by Smith and Heshusius, dismiss the implications of noting such fundamental differences with a statement favouring pragmatism: "Epistemological purity doesn't get research done" (p. 21). In psychological terms, Miles and Huberman compromise: they recognize that an essential difference exists but fail to address it. Other researchers have also taken this stand as evidenced by their advocacy of certain types of triangulation (Goodwin & Goodwin, 1984; Mitchell, 1986; Swanson-

Laufman, 1986). Triangulation combines multiple sources of data, investigators, hypotheses or several different methods. If triangulation comprises qualitative and quantitative methods in one study, they would have to be used disjunctively and not conjunctively to maintain epistemological purism (Howe, 1985).

It is of interest that certain qualitative researchers do not acknowledge the essential differences (Le Compte & Goetz, 1982), or do acknowledge them and do not act on them (Miles & Huberman, 1984). Smith and Heshusius (1986), in contrast, stay close to the historical, philosophical thought that helped spawn qualitative methods. The debate as to who is right or wrong is not simple. If one believes in the "mind involvement of a constituted reality" (Smith, 1983, p. 9), then "truth" would be viewed as agreement among people in any given time and space; "objectivity" would be social agreement with values and facts integrated into the research process. The very act of believing in the philosophical tenets of idealism should make it impossible to engage simultaneously in research based in positivism; yet, what has historically been known as truth may not be the same truth as is currently understood. Dismissing Smith and Heshusius's position and getting on with it, as Miles and Huberman advocate is not sufficient; reasons for these differences must be ascertained. Have the perceptions of truth, reality and meaning changed? Is rigid adherence to philosophical thought advisable or even possible? The differences do exist and there is presently a research trend toward "epistemological ecumenism" (Miles & Huberman, p. 20). This is concretely expressed when terms such as validity and reliability are applied to qualitative methods.

Rigour in Qualitative Methods

One of the most pressing problems in qualitative research is to develop enquiry criteria and procedures that will sanction it as a viable research process with a common language structure, clearly defined purposes, methods and analyses. In short, to find criteria and procedures for qualitative methods unlike, but as strong as, those used by quantitative researchers, (i.e., measurement theory, sampling rules and statistical analysis).

Compounding the problem is the great diversity of methods and procedures loosely associated with qualitative enquiry that stems from a number of disciplines, each with its own rules and goals. For example, ethnography, symbolic interactionism and grounded theory were each primarily developed in different fields (Bodgan & Bilken, 1982). Phenomenology, also grouped under qualitative enquiry, began with the work of Kant when, at the end of the eighteenth century, he described man as "knowing". Difficulties arise when researchers cross disciplines and force their view of the world on others rather than working collaboratively for a common purpose. At the

same time, researchers open to other "ways of knowing" have developed very creative, commendable methods of qualitative enquiry that have proven to be meaningful to other researchers and society at large (Glaser & Strauss, 1960; Rogers, 1985).

Assessment framework

Sandelowski (1986) acknowledges these problems and describes a broad framework (referring to the work of Guba and Lincoln) to assess rigour. The four criteria in the framework are: credibility, fittingness, auditability and confirmability.

Credibility occurs when the people (participants) involved in the research read descriptions and immediately recognize the lived experiences to be their own; the description may be of another participant, but it is believed it could be that of the reader himself. As well, credibility is increased when the researchers describe their relationships with the participants: behaviours, impressions and experiences. The relationship need not be positive to be informative, as was the case with the records of Malinowski; Kirk and Miller (1986) observed that if Malinowski's personal diary is to be believed he was a very neurotic and bitchy man.

Fittingness means that the "finding can 'fit' into contexts outside the study situation" (Kirk and Miller, 1986, p. 32). The researcher should consider how the sampling was done, and obtain data that are as representative as possible for the experience under investigation. Fittingness would also refer to whether the results were seen as meaningful to participants not involved in the study.

Auditability refers to the ability of another researcher to follow the thinking, decisions and methods used by the original researcher. The second researcher would arrive at similar results, but not contradictory conclusions. For auditability to occur, the original researcher should painstakingly document the entire research process through the use of memos, diaries or field notes, mechanical recordings, minutes of meetings, letters, etc. A classic example of how auditability was achieved but rendered ineffective occurred when Mead and Freeman studied adolescent sexual behaviour in Samoan society. Mead reported that virginity was nominally important whereas Freeman said virginity was extremely important (Kirk & Miller, 1986). The difference in the findings was not attributed to the methods of data collection but to the gender differences of the researchers: young Samoan females were more revealing to a female researcher than to a male researcher.

Confirmability is a criterion for neutrality; it is achieved when credibility, auditability and fittingness are established. Again, if the participants and

others observe that the findings are meaningful to their lived experiences, confirmability is achieved. Glaser and Strauss (1966) indirectly address confirmability when they observe that qualitative analysis is often the end product of research because the unknown becomes vividly known, saturation of knowledge frequently shifts interest to another phenomena and laymen profit from the findings.

The important questions are not how reliable and valid are the methods and results of a study, but how credible, fitting, auditable and confirmable overall are the findings? Similar terms have been used by Strauss and Glaser (1966) and May (1985) in direct reference to grounded theory. In qualitative research, the researcher is fully dependent on the participant's evaluation of the findings; subjectivity is recognized and accepted as part of the research process. The lived experience of the participant is real.

Challenges for Qualitative Researchers

(1) Develop and use rules, terms and procedures to describe the qualitative research process accurately.

(2) Ensure that participants are actively involved in all phases of the research project, including being present at dissemination of the findings through presentations and are informed of publications.

(3) Understand the purposes and implications of using terms such as validity and reliability.

(4) Tolerate uncertainty and confusion (particularly when pitted against an articulate positivist!) as a new language to describe the relevance of qualitative enquiry emerges.

(5) Recall that there is an essential difference between qualitative and quantitative methods of enquiry and that if both are to be mixed, the researcher must provide a sound rationale.

REFERENCES

- Allender, J.S. (1986). Educational research: A personal and social process. *Review of Educational Research*, 56(2), 173-193.
- Atwood, J.B., & Hinds, P. (1986). Heuristic heresay: Application of reliability and validity criteria to products of grounded theory. *Western Journal of Nursing Research*, 8(2), 135-154.
- Bogdan, R.C., & Bilken, S.K. (1982). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn and Bacon, Inc.
- Duffy, M. (1985). Designing nursing research: The qualitative-quantitative debate. *Journal of Advanced Nursing*, 10(3), 225-232.
- Glaser, B.G., & Strauss, A.L. (1966). The purpose and credibility of qualitative research. *Nursing Research*, 15(1), 56-61.
- Goodwin, L.D., & Goodwin, W.L. (1984). Qualitative vs quantitative research or qualitative and quantitative research? *Nursing Research*, 33(6), 378-380.
- Howe, K. (1985). Two dogmas of educational research. *Educational Researcher*, 14, 10-18.
- Kirk, J., & Miller, M.L. (1986). *Reliability and validity in qualitative research*. London: Sage Publications.
- Knapp, T.R. (1985). Validity, reliability and neither. *Nursing Research*, 34(3), 189-192.
- Le Compte, M.D., & Goetz, J.P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research*, 52(1), 31-60.
- Leininger, M.M. Ed. (1985). *Qualitative research methods in nursing*. Toronto: Grune and Stratton, Inc.
- May, K.A. (1986). Writing and evaluating the grounded theory research report. In Chenitz, W.C. and Swanson, J.M. (1986). *From practice to grounded theory*. Don Mills: Addison-Wesley Pub. Co.
- Miles, M.B., & Huberman, A.M., (1984). Drawing valid meaning from qualitative data: toward a shared craft. *Educational Researcher*, 13(5), 20-28.
- Mitchell, E.S. (1986). Multiple triangulation: A methodology for nursing science. *Advances in Nursing Science*, 8(3), 18-26.
- Rodgers, C. (1985). Toward a more human science of the person. *Journal of Humanistic Psychology*, 25(4), 7-24.
- Sandelowski, M. (1986). The problem of rigour in qualitative research. *Advances in Nursing Science*, 8(3), 27-37.
- Smith, J.K. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. *Educational Researcher*, 12(3), 6-13.
- Smith, J.K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquires. *Educational Researcher*, 15(1), 4-12.
- Swanson-Laufmann, K.M. (1986). A combined qualitative methodology for nursing research. *Advances in Nursing Science*, 8(3), 58-69.
- Swanson, J.M., & Chenitz, W.C. (1982). Why qualitative research in nursing? *Nursing Outlook*, 30(4), 241-245.
- Stern, P.N., Allen, L.M., & Moxley, P.A. (1984). Qualitative research: The nurse as grounded theorist. *Health Care for Women International*, 5-6(5), 371-385.

Support of this article came from an Alberta Foundation of Nursing Traineeship in the form of release time to study research methods.

RÉSUMÉ

Fiabilité et Validité:

Termes mal appropriés à la recherche qualitative

Dans ce papier, on demande pourquoi les règles qui gèrent la fiabilité et la validité ne pourraient être appliquées à l'enquête qualitative? Nous discuterons de deux questions fondamentales: (a) les différences essentielles qui existent entre la recherche qualitative et la recherche quantitative, et (b) la rigueur obtenue par la voie d'autres critères, tels que: crédibilité, aptitude, auditabilité et confirmabilité. Le défi, pour les chercheurs en sciences infirmières, consiste à développer des termes, règles et procédés propres à la recherche qualitative, plutôt que de superposer des règles et termes quantitatifs aux méthodes qualitatives.