External Validity, Generalizability, and Knowledge Utilization

ARTICLE *in* JOURNAL OF NURSING SCHOLARSHIP · FEBRUARY 2004

Impact Factor: 1.64 · DOI: 10.1111/j.1547-5069.2004.04006.x · Source: PubMed

1 AUTHOR:

Linda Ferguson
University of Saskatchewan
41 PUBLICATIONS **435** CITATIONS

SEE PROFILE

External Validity, Generalizability, and Knowledge Utilization

Linda Ferguson

Purpose: *To examine the concepts of external validity and generalizability, and explore strategies to strengthen generalizability of research findings, because of increasing demands for knowledge utilization in an evidence-based practice environment.*

Framework: *The concepts of external validity and generalizability are examined, considering theoretical aspects of external validity and conflicting demands for internal validity in research designs. Methodological approaches for controlling threats to external validity and strategies to enhance external validity and generalizability of findings are discussed.*

Conclusions: *Generalizability of findings is not assured even if internal validity of a research study is addressed effectively through design. Strict controls to ensure internal validity can compromise generalizability. Researchers can and should use a variety of strategies to address issues of external validity and enhance generalizability of findings. Enhanced external validity and assessment of generalizability of findings can facilitate more appropriate use of research findings.*

JOURNAL OF NURSING SCHOLARSHIP, 2004; 36:1, 16-22.

[Key words: external validity, generalizability, research utilization, knowledge utilization]

* * *

Although nursing research has reduced the gap between theory and practice, the gap persists (Chinn & Kramer, 1999; Schmitt, 1999). Practitioners often report that the findings of research studies are difficult to use in practice, or are irrelevant to their practice issues, or are inaccessible to practitioners (Estabrooks, 1998; Huberman, 1995; Lomas, 2000). Researchers bemoan the lack of use of their study results in practice, despite efforts to disseminate findings (Huberman, 1995; Landry, Amara, & Lamari, 2001; Lomas, 2000; Rich, 1991). This disparity between nursing researchers and practitioners is evidence of the continuing challenge to link theory, research, and practice.

The main premise of this article is that an increased focus on external validity of studies and generalizability of findings could increase the use of knowledge generated in those studies. External validity can be viewed as the conceptual as well as the effective link between knowledge generation and knowledge utilization. Researchers should design and report their studies with reasonable generalization of findings as a goal.

The topic of external validity is largely absent in nursing literature. A search of CINAHL between 1982 and 2002 revealed no nursing articles and only three editorials dealing with external validity or generalizability as the main focus (Hayes, 1998; Hegyvary, 2002; Schmitt, 1999). Researchers need specific strategies to address external validity as a critical element in research design and in utilization of research-generated knowledge in clinical practice and policy development.

Validity of Research Findings

Cook and Campbell (1979) described validity as the best available approximation to the truth or falsity of propositions, including propositions about causation. Although this assertion indicates a controversial post-positivist approach to the question of truth, specifically that one "truth" is achievable, it does provide the basis for applying research findings to other populations, times, or settings. Campbell and Stanley (1963) delineated two types of validity: internal validity as a characteristic of the experimental treatment effect, and external validity that provides the basis for generalizability to other populations, settings, and times. Campbell and Stanley described internal validity as "the basic minimum without which any experiment is virtually uninterpretable" (p. 5).

Linda Ferguson, RN, MN, *Mu Sigma*, Professor, College of Nursing, University of Saskatchewan, Saskatoon, Saskatchewan, Canada. Correspondence to Ms. Ferguson, College of Nursing, University of Saskatchewan, 107 Wiggins Road, Saskatoon, Saskatchewan S7N 5E5, Canada. E-mail: linda.ferguson@usask.ca

Accepted for publication April 14, 2003.

Internal validity refers to the confidence with which one can make statements about relationships between variables, based on the forms in which the variables were manipulated or measured (Cook & Campbell, 1979). Shadish, Cook, and Campbell (2002) emphasized that internal validity is an experimental concept about the relationship of research operations, irrespective of what they theoretically represent. Choice of design can control for some of the threats to internal validity and give researchers greater confidence that the effects in the study were attributable to the experimental or treatment variable and not to other plausible explanations for the observed effects. Statistical conclusion validity, which also contributes to internal validity, shows inferences made about cause and effect variables, based on statistical evidence (Cook & Campbell, 1979).

Internal validity is important in research design, and historically it has been and will continue to be the main focus of researchers. Establishing internal validity is a logical process and is planned a priori in the design of the study (Shadish et al., 2002; Slack & Draugalis, 2001). Unfortunately, an inverse relationship between internal validity and external validity exists: As researchers plan for control of extraneous variables that could confound the findings of causal relationship, they limit the external validity and generalizability of those findings (Christensen, 2001; Shadish et al., 2002).

External Validity

First described by Campbell and Stanley (1963), external validity pertains to the generalizability of the treatment effect to other populations, settings, treatment variables, or measurement variables. Christenson (1985) elaborated on this definition by discussing the generalization of treatment effects to and among other populations, settings, and times. Shadish and colleagues (2002) defined external validity as the validity of inferences about whether the identified causal relationship is maintained over variations in person, settings, time, or treatment variables. Researchers often attend to issues of external validity, if at all, only after they have carefully considered the internal validity of the study. Unfortunately, controlling for threats to internal validity often results in reduced external validity of those findings.

Cook and Campbell used the terms "generalize to" and "generalize across" to distinguish between situations in which the sample is randomly selected from the target population, and thus findings can be generalized to that population with confidence. Generalizing across populations includes applying findings to populations, settings, or times that were not represented in the sample. In most field research situations, the sample is not representative of the target population through randomization, and thus the findings pertain only to the sample of the study. When findings are generalizable to and across populations, the process is inferential and broader statements of application are made, based on limited findings shown within the research study (Christensen, 2001; Slack & Draugalis, 2001).

External validity is frequently associated with the term "generalizability," often being used interchangeably with it. Generalization is the process of using particular data to infer a general statement that has applicability to other people, settings, or times. Generalizability refers to the characteristic of applicability in other situations. Shadish and colleagues (2002) extended the concept to include treatment and outcomes, and they differentiated two types of generalization, one related to conceptual constructs and theory, and the other related to the causal relationship. External validity and generalizability are based on the assumption of uniformities in nature and the existence of natural laws. If such uniformities exist, then valid research results can be generalized to and across different people, settings, and times (Christensen, 2001).

External validity is a function of the researcher and the design of the research. Generalizability is a function of both the researcher and the user. This process of generalization is inferential based on knowledge of the representativeness of the sample to the referent or target population, including formal probability and nonprobability sampling (Morgan & Harmon, 1999). Representativeness of the sample theoretically allows for generalization of the results of the study to the target population (Christensen, 2001). For knowledge utilization, the user must determine the relevance of the findings for the proposed population, and whether the findings are generalizable to the user's setting and population.

Threats to external validity. In a practice environment that is increasingly more evidence-based, researchers need to consider the factors that threaten the external validity of their studies (Rich & Oh, 2000). Campbell and Stanley (1963) identified four factors jeopardizing external validity or representativeness: (a) interactional effects of testing, (b) interactional effects of selection biases with experimental variables, (c) reactive effects of experimental arrangements, and (d) multiple treatment interference. External validity is a methodologic concept that researchers address by choosing a design with attention to both internal and external validity. Cook and Campbell (1979) extended the discussion of external validity to include construct validity.

Interaction of selection and treatment (Campbell & Stanley, 1963; Cook & Campbell, 1979), also called interaction of "causal relationship" with units (Shadish et al., 2002), is a key threat to external validity. Cook and Campbell (1979) described the importance of sampling to identify the target population, select samples to represent that population, and recognize when samples are selected on the basis of convenience or experimentally accessible populations (Christensen, 2001; Morgan & Harmon, 1999). Samples might include people or aggregate units such as nursing units or organizations (Shadish et al., 2002).

Random selection of participants or units likely assures a representative sample of that target population, and thus findings are generalizable to the experimentally accessible population. Christensen (2001) suggested that deliberate sampling for heterogeneity or stratified random sampling might be initiated to facilitate generalization. As Shadish

and colleagues (2002) discussed, random sampling is rarely feasible in field studies; purposive sampling is most often used to select samples that are diverse on specific variables. Researchers enhance the generalizability of their findings by assessing the representativeness of the sample statistically, because limitations in representativeness complicate inferences from the sample to the target population. The issue then becomes one of determining whether researchers and decision makers can generalize findings across populations not represented explicitly by the sample.

Researchers might have recruited people or units with particular characteristics such as reading ability or socioeconomic status from the targeted population, such that participants are not representative of that population, especially if participation in the experimental treatment is arduous or difficult. Cook and Campbell (1979) suggested that researchers design their studies to facilitate retention of participants and assess mortality or dropout rates in the treatment groups carefully to determine if people with specific characteristics have withdrawn. Researchers should analyze group homogeneity and representativeness at the beginning and end of the study. Christensen (2001) referred to this issue as "population validity."

Merely addressing issues of sample selection does not assure external validity; other threats to external validity must also be addressed. Interaction of setting and treatment becomes an issue if the setting of the experimental treatment is artificial (Cook & Campbell, 1979). Christensen (2001) defined "ecological validity" as the ability to generalize findings across settings or from one set of environmental conditions to another. Christensen identified several issues within ecological validity, including the threat of multiple treatment interference (measurement effect), in which the effect of participation in one aspect of the treatment affects the outcome findings in other treatment conditions, also known as sequencing effect. The effect of participating in an experiment, the Hawthorne effect, or novelty or disruption factor might also affect ecological validity by an effect that is the result of conditions other than the experimental treatment. Christensen also addressed the effect of pretesting as a threat to ecological validity. Varying the setting and context, especially using multiple sites, reduces the threat that findings are relevant only in the experimental setting.

The third threat to external validity involves the interaction of history and treatment, raising the question of applicability of findings to other periods of time. Christensen (2001) addressed this issue as temporal validity, raising questions about the timing and frequency of data collection in relation to seasonal or cyclical variations. Replication studies might be focused on the effects of single events, thus increasing external validity of the findings and increasing the likelihood that the findings are relevant in the future.

Shadish and colleagues (2002) indicated that interaction of causal relationships (treatments) with both treatment variations and outcome variations affects external validity. The causal relationship that has been shown does not hold for all levels of treatment or possible outcomes, and it must be generalized with caution for those variations. Although effect size might not be constant across different units or groups of participants, the direction of the causal relationship is usually constant, and thus allows for generalization (Shadish et al., 2002).

Construct Validity

To extend beyond Campbell and Stanley's focus on definition and measurement of variables and representativeness of sample, Cook and Campbell (1979) associated construct validity with external validity. Construct validity is the relationship of the operational definitions of variables to their conceptualizations. Construct validity indicates that the operations that are meant to represent particular variables are in fact representative and exclusive. Construct validity indicates that researchers in specific fields generally accept conceptualizations of variables that are theory-based to have a particular meaning. Specifically, research measures adequately and exclusively represent their referent constructs (Bear, 1990).

Both external validity and construct validity relate to generalizability of findings (Cook & Campbell, 1979). Construct validity refers to the fit between the cause-and-effect operational definitions and their referent constructs (Bear, 1990). Constructs and theoretical propositions should be clearly explicated before designing and conducting the study. Construct validity has a convergent component, in that various instruments used to measure a construct are consistent in their assessment of that construct, and a divergent component, in that measures also indicate distinction between related but conceptually different concepts. Shadish and colleagues (2002) indicated that clarity in the construct(s) under consideration is important to make inferences from the research operations and measures to their higher-order constructs, to generalize the treatment effects to other populations, and to integrate findings generated by the research with other theories and bodies of knowledge.

Goodwin and Goodwin (1991) described construct validity as an "umbrella" concept that encompasses content, concurrent, convergent, divergent, and predictive validity, and as the theoretical concept underlying the hypothesis and indicating the type of data collected. Construct validity can be shown qualitatively by elaborating on theoretical conceptions and quantitatively by correlating different measures of the construct. Generalizability of effects to other settings, time, and populations is enhanced by the relationship of the conceptual underpinnings of the research to accepted knowledge of the discipline (Hegyvary, 2002). Without these theoretical underpinnings, findings would be limited to the research operations as the context and would be difficult to generalize to other situations. As an example, if a researcher uses a particular measurement tool to collect data on a specific concept, such as hope, the relationship between the concept and the tool should be evident in the literature.

Cook and Campbell (1979) and Shadish and colleagues (2002) indicated that the manner in which the study

is conducted, including participant eagerness to "do it right," experimenter expectancy of outcome, and novelty disruption can negatively affect construct validity. Single measures of theoretical constructs might also result in under-representation, known as mono-operational bias. The possibility of inadequately representing a particular construct through the experimental design and operation jeopardizes external validity and limits generalizability. Cook and Campbell (1979) used the term convergence to describe the congruence in measures or manipulations of the same construct, and divergence to indicate distinction between measures of unrelated, albeit similar, constructs (p. 61). Lack of construct validity and external validity limits knowledge utilization.

Knowledge Utilization

Knowledge utilization, the overarching term that includes research utilization and evidence-based practice, pertains to the use of knowledge generated through research for policy and practice decisions (Rich, 1991). It includes research, scholarly practice, and programmatic interventions aimed at increasing the use of knowledge to solve human problems (Backer, 1991). With increasing emphasis on evidence-based practice (Estabrooks, 1998; Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996) and evidence-based policy (Backer, 1991; Beyer, 1997; Beyer & Trice, 1982; Butcher, 1998; Huberman, 1995; Rich & Oh, 2000), researchers, practitioners, and policymakers are searching for ways to increase use of research findings. Increasing the transfer of findings into practice and policy development is the goal (Campbell & Stanley, 1963). This focus on knowledge utilization is gaining importance; however, knowledge utilization remains limited (Lomas, 2000).

The failure to use research findings in practice and policy is complex. From policymaker and practitioner perspectives, these factors include the relevance of knowledge generated by research, the time that research takes, the failure to address practice issues comprehensively, and limited applied research. Policy makers and practitioners describe difficulties accessing knowledge in a timely manner, accessing all studies relevant to the topic, difficulties in reconciling the contradictory outcomes of studies, lack of replication studies, and the short times for both practice and policy decisions. Decision makers indicate preference for context-relevant research (applied research), and they report lack of understanding between researchers and practitioners (Beyer, 1997; Lomas, 2000; Rich & Oh, 2000).

Estabrooks (1999) and Beyer (1997) described three distinct ways of using research knowledge for policy development. Instrumental use, the direct use of research findings to support practice or policy development, is limited, and it varies depending on the field and context of practice or policy. Conceptual use is more frequent; research findings influence attitudes and beliefs, but no attributable actions are taken. Use of research findings might also be symbolic or persuasive; research findings are used to support previous decisions. In part, failure to use knowledge generated through research might relate to how one regards its use. Because most research findings are used conceptually or symbolically, but most knowledge utilization studies are focused on instrumental use (Beyers & Trice, 1982), use of research knowledge is deemed limited. In contrast, Landry and associates (2001) described a six-step cumulative process in knowledge utilization that included conceptual use, and they demonstrated much higher levels of use.

Regardless of how knowledge use is measured, many issues remain regarding the nature of evidence needed for decision making and the design of the research. Lomas (2000) used the term "two communities" to describe the distance between researchers and practitioners, reflecting researchers' failure to understand the complexities of knowledge use in practice, and practitioners' failure to understand complexities of the research process. A bridge to better understanding between the two communities might include researchers' increased focus on external validity and partnerships with practitioners to better address their practice and policy issues (Beyer, 1997; Rich, 1991; Rich & Oh, 2000).

Researchers have generally approached their studies with a strong focus on the quality of the knowledge generated from those studies and a lesser focus on the relevance of that knowledge for practitioners and other users (Shadish et al., 2002). This continuing commitment to internal validity is understandable and commendable. However, if researchers increased their focus on the external validity of their findings, this emphasis could assist health care practitioners and policy makers in use of this knowledge.

Strategies to Enhance External Validity or Generalizability

Researchers can increase the external validity and generalizability of their findings and the use of the findings in many ways. Researchers will continue to focus on internal validity, while recognizing that strong internal validity introduces threats to external validity. Careful analysis of these threats to external validity will indicate limits to generalizability. In openly acknowledging these limitations in their reports, researchers assist users to make informed decisions about generalizing findings to new settings or populations.

Research design can contribute to greater generalizability of results. Threats to external validity such as treatment and participant interactions can be addressed methodologically through sample selection. Random selection and random assignment of participants, although not always feasible, increase the generalizability of findings to the target population, and possibly to other populations. Threats such as interactions of treatment and setting, or of treatment and history, can be addressed methodologically by introducing multisite aspects to the study, by addressing timing of studies such that history varies, or through replication (Fahs, Morgan, & Kalman, 2003; Hayes, 1998). Researchers could acknowledge historical events occurring during the course

of the study, and, if possible, address these threats through data collection in a variety of circumstances. Researchers also assist users to make informed generalization of findings by assessing the results of multiple treatments, variation in treatment intensity, or pretesting on the treatment effect. Careful analysis of the generalizability of their findings based on their methodological design is beneficial to practitioners, policymakers, and future researchers.

A key factor in how research findings and research-generated knowledge are used is the manner in which researchers present their findings and the language that is used (Rich, 1991). Dissemination of research findings in appropriate journals facilitates practitioner access and utilization. Presentation of findings in an understandable manner, such that external validity and limitations are clearly evident, is also important to facilitate knowledge utilization. For policy developers, this may be in the form of presentations, executive summaries, or fact sheets, and, for practitioners, in the form of reasonable recommendations for generalization. Although this practice does not change the external validity of the study, it does facilitate the user's determination of whether the findings are generalizable to one's own population and setting. In general, research findings must be targeted to users in an appropriate and relevant format.

In their reports and manuscripts, researchers should discuss sample selection, including the number and demographic characteristics of units or people who declined to participate, and a statistical description of mortality. Use of random selection and the sample selection from the target population instead of merely from the experimentally accessible population should be clearly evident in the report (Morgan & Harmon, 1999). The report should also include effect size and level of significance, allowing readers to identify clinical significance versus statistical significance (LeFort, 1993). Researchers should report sufficient detail to allow for replication of the study to determine if the same results are obtained in other circumstances.

Reports of studies should be in sufficient detail (including data and results) to allow readers to logically assess the effects to both internal and external validity (Slack & Draugalis, 2001) and to determine generalizability. The methods section should include the study design and procedures taken to control specific threats to internal validity. The results section should include data relevant to establishing internal validity, such as interrater reliabilities or comparison group homogeneity. The discussion section should provide the researcher's assessment of the influence of bias on study results. If specific threats to internal validity are evident, readers should be able to trace measures taken to address them. Important issues related to experimental mortality, blinding, randomization, adherence to protocols, and placebo effects must be assessed in relation to design, results, and discussion of possible effects.

Researchers should be cautious in reporting the generalizability of their research findings and the limitations of the study. Generalizing beyond the sample of the study should be based on the size and representativeness of the sample, target population, analytic methods, setting, and theoretical constructs (Hegyvary, 2002; Shadish et al., 2002). Findings from smaller studies with limited (or no) generalizability should be put into the context of the literature on the same topic, adding to extant disciplinary knowledge, but not being directly applied (Hegyvary). Such contributions can add to theoretical knowledge of a phenomenon, leading to future research in practice settings.

Shadish and colleagues (2002) proposed five principles to guide researchers in their generalizations of research findings. In making statements about generalizability, researchers should engage in the following process: (a) assess the apparent similarities between research operations and the prototypical characteristics of the target of generalization (surface similarity); (b) rule out irrelevancies by identifying attributes of person, settings, treatments, and outcome measures that do not affect generalization (ruling out irrelevancies); (c) identify those features of people, settings, treatments, or outcomes of the study that limit generalization (discriminating); (d) generalizing causal relationships (interpolating and extrapolating); and (e) developing and testing explanatory theories about the target of generalization (causal explanation). Consideration of these principles can help researchers to make reasonable statements about generalizability of their findings. Practitioners and policymakers can use the same principles to determine applicability of reported findings to their own situations.

Key determinants of generalizability are the representativeness of the sample and the setting of the study. Researchers should be cautious in discussion of implications of or application of findings from studies that have limited external validity. In many studies, particularly qualitative studies, small samples might preclude generalizability of findings to other populations; however, integrating the findings with other similar research studies reported in the literature can be useful for positioning the study within the knowledge of the discipline (Hegyvary, 2002). In this way, smaller studies can contribute to theory that can be applied and evaluated in other settings and populations.

Lynch (1999), Winer (1999), Bear (1990), and Huberman (1995) emphasized the importance of controlling for threats to construct validity and generalizing theory. To generalize theory to other populations or to add to existing knowledge of the discipline, research concepts, hypotheses, operational definitions and measures must be based within the theoretical framework (Chinn & Kramer, 1999). Evaluation research on the application of theory to and across populations should be conducted and reported, especially for policy development (Rich, 1991; Whittemore & Grey, 2002) and for international application (Hayes, 1998). Research findings might generalize to and across other populations and settings in the form of theory, and thus add to the knowledge of the discipline, but this generalization is possible only if the research operations are strongly based in theory (Bear, 1990; Shadish et al., 2002).

Winer (1999) and Lynch (1999) indicated the need for further studies to demonstrate external validity of the findings

of a study, suggesting that the original research might have limited external validity. Replication studies and evaluation studies are key to assessing the proposed external validity of the original research findings, especially if the results have been or are expected to be incorporated into health or social policy (Fahs et al., 2003; Hayes, 1998; Lynch, 1999). Replication studies can reduce the threats to external validity, thus improving generalizability. Finding similar results in a variety of settings, populations, and times increases confidence in the validity of those findings for other circumstances. Replication studies can also correct limitations and add evidence to protect against error, thus increasing the strength of the evidence for specific practice and reconciling conflicting findings (Fahs et al., 2003).

Such a focus on follow-up research to determine external validity is an aspect of programmatic research that includes studies through the full phases of exploration and experimentation, as well as evaluation of implementation (Whittemore & Grey, 2002). This focus on programs of study also indicates a high level of tacit knowledge in the area such that the researcher who has controlled for moderating factors can have knowledge to support re-contextualization of the findings in new populations and settings.

Programs of research should show a progression through concept clarification and theory development to theory testing (Whittemore & Grey, 2002). Adding evaluation studies of research findings implemented in practice could enhance the program of research and show the generalizability of findings in other settings. Using the National Institutes of Health (NIH) framework, Whittemore and Grey proposed that investigating nursing treatments and interventions progress through five phases to facilitate external validity of study results, including concept clarification, pilot testing, clinical efficacy trials (randomized controlled trials), clinical effectiveness testing, and wide-scale implementation studies. Internal validity would be the primary focus of earlier phases, but after efficacy is clearly demonstrated, greater emphasis can confidently be placed on external validity and generalizability of findings to other populations (Whittemore & Grey, 2002).

Use of meta-analysis and meta-synthesis to increase sample size and demonstrate applicability of findings in multiple settings facilitates incorporating findings into practice and policy development (Beck, 1999; Onyskiw, 1996). This approach includes analyzing research reports to determine the validity of findings, and consolidating them such that treatment effects are evident over multiple populations, settings, and times, thus increasing the external validity of the overall results. Meta-analysis is particularly useful for practitioners and policy makers who may not have either the expertise or time to do a thorough search for relevant reports or analysis of component studies independently. The Cochrane Collaboration with studies of health (Clarke & Oxman, 2002) and the Campbell Collaboration with studies of education are examples of organization that undertake such reviews in a very systematic and organized manner and make their reviews readily available. To increase their effectiveness, these reviewers could report those subpopulations for whom no effect or a negative effect was found, such that generally applicable findings would not be inappropriately generalized.

Researchers have a responsibility to report their studies fully including critical data, to facilitate meta-analysis, as well as generalizability and application (Beck, 1999). The theoretical framework of the study is an essential element for contextualizing the findings of the study. Accuracy in reporting data and statistical values is critical in supporting subsequent meta-analysis, consolidating contradictory findings, increasing power of smaller studies, and generalizing findings. Systematic reviews include comprehensive analysis of current research on the topic, such that practitioners and policy makers can make decisions based on findings with greater confidence.

As an important step in enhancing external validity, funding agencies are encouraging partnerships among researchers, practitioners, and policy makers to effectively bridge the gap between "communities" and to facilitate knowledge utilization (Beyer, 1997; Rich, 1991). Ideally, community partners participate in research from early stages of development, and provide for linkages to encourage researchers to address real-life problems and decision makers to use findings of the research, especially as research programs progress to studies of clinical effectiveness or applied research (Huberman, 1995; Rich & Oh, 2000). The benefits of such partnerships are significant, particularly in the area of generalizability.

Conclusions

External validity is critical for generalizing and applying research findings. The veracity of findings of research studies is important, and both researchers and knowledge users need to be able to rely on the findings. Researchers can enhance the external validity of their studies while addressing issues of internal validity by ensuring representativeness in the sample and controlling moderating factors that jeopardize external validity. Researchers can enhance assessment of external validity and knowledge utilization through appropriately reporting their findings, developing programs of research often in partnerships with other researchers with complementary research skills, facilitating meta-analysis of studies, and developing research partnerships with practitioners and policy makers to address relevant issues in a timely manner. Researchers should consider external validity as a key to knowledge utilization. External validity is an essential link between knowledge generation and knowledge utilization.

References

Backer, T.E. (1991). Knowledge utilization: The third wave. **Knowledge: Creation, Diffusion, Utilization, 12**(3), 225-240.

Bear, M. (1990). Theoretical issues in measurement. In L. Moody (Ed.), **Advancing nursing science through research**. Newbury Park, CA: Sage.

Beck, C.T. (1999). Facilitating the work of a meta-analyst. **Research in Nursing & Health, 22**, 523-530.

Beyer, J.M. (1997). Research utilization: Bridging a cultural gap between communities. **Journal of Management Inquiry, 6**(1), 17-22.

Beyer, J.M., & Trice, H.M. (1982). The utilization process: A conceptual framework and synthesis of empirical findings. **Administrative Science Quarterly, 27**, 591-622.

Butcher, R.B. (1998). Foundations for evidence-based decision making. In National Forum on Health (Canada), **Evidence and information: Making decisions**. Saint-Foy, Quebec: Editions MultiMondes.

Campbell, D.T., & Stanley, J.C. (1963). **Experimental and quasi-experimental designs for research**. Boston: Houghton Mifflin.

Chinn, P.L., & Kramer, M.K. (1999). **Theory and nursing: Integrated knowledge development** (5th ed.). St. Louis, MO: Mosby.

Christensen, L.B. (2001). **Experimental methodology** (8th ed.). Boston: Allyn & Bacon.

Clarke, M., & Oxman, A.D. (Eds.). (2002). Cochrane reviewers handbook 4.1.5. [updated April 2002]. In **The Cochrane Library, Issue 2**. Oxford: Update Software.

Cook, T.D., & Campbell, D.T. (1979). **Quasi-experimental experimentation: Designs and analysis for field settings**. Boston: Houghton Mifflin.

Estabrooks, C.A. (1998). Will evidence-based nursing practice make practice perfect? **Canadian Journal of Nursing Research, 30**(1), 15-36.

Estabrooks, C.A. (1999). The conceptual structure of research utilization. **Research in Nursing and Health, 22**, 203-216.

Fahs, P.S., Morgan, L.L., & Kalman, M. (2003). A call for replication. **Journal of Nursing Scholarship, 35**, 67-72.

Goodwin, L.D., & Goodwin, W.L. (1991). Estimating construct validity. **Research in Nursing & Health, 14**, 235-243.

Hayes, P. (1998). The issue of generalizability. **Clinical Nursing Research, 7**(3), 227.

Hegyvary, S.T. (2002). Knowledge for a diverse world (editorial). **Journal of Nursing Scholarship, 34**, 205.

Huberman, M. (1995). Research utilization: The state of the art. **Knowledge and Policy, 7**(4), 13-34.

Landry, R., Amara, N., & Lamari, M. (2001). Utilization of social science research knowledge in Canada. **Research Policy, 30**, 333-348.

LeFort, S.M. (1993). The statistical versus clinical significance debate. **Journal of Nursing Scholarship, 25**, 57-61.

Lomas, J. (2000). Using 'Linkage and Exchange' to move research into policy at a Canadian foundation. **Health Affairs, 19**(3), 236-240.

Lynch, J.G. (1999). Theory and external validity. **Journal of the Academy of Marketing Science, 27**(3), 367-376.

Morgan, G.A., & Harmon, R.J. (1999). Sampling and external validity. **Journal of the American Academy of Child and Adolescent Psychiatry, 38**(18), 1051-1053.

Onyskiw, J.E. (1996). The meta-analytic approach to research integration. **Canadian Journal of Nursing Research, 28**(3), 69-85.

Rich, R.F. (1991). Knowledge creation, diffusion, and utilization. **Knowledge: Creation, Diffusion, Utilization, 12**(1), 319-337.

Rich, R.F., & Oh, C.H. (2000). Rationality and the use of information in policy decisions: A search for alternatives. **Science Communication, 22**(2), 173-211.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., & Richardson, W.S. (1996). Evidence based medicine: What it is and what it isn't. **British Journal of Medicine, 312**, 71-72.

Schmitt, M.H. (1999). Closing the gap between research and practice: Strategies to enhance research utilization (editorial). **Research in Nursing & Health, 22**, 433-434.

Slack, M.K., & Draugalis, J.R. (2001). Establishing the internal and external validity of experimental studies. **American Journal of Health-Systems Pharmacists, 58**(22), 2173-2181.

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

Whittemore, R., & Grey, M. (2002). The systematic development of nursing interventions. **Journal of Nursing Scholarship, 34**, 115-120.

Winer, R.S. (1999). Experimentation in the 21st century: The importance of external validity. **Journal of the Academy of Marketing Science, 27**(3), 349-358.