Generalization in quantitative and qualitative research: Myths and strategies

Denise F. Polit [a,b,*], Cheryl Tatano Beck [c]

[a] Humanalysis, Inc., 75 Clinton Street, Saratoga Springs, NY 12866, United States
[b] Research Centre for Clinical and Community Practice Innovation, Griffith University School of Nursing, Gold Coast, Australia
[c] University of Connecticut School of Nursing, Storrs, CT, United States

ABSTRACT

Generalization, which is an act of reasoning that involves drawing broad inferences from particular observations, is widely-acknowledged as a quality standard in quantitative research, but is more controversial in qualitative research. The goal of most qualitative studies is not to generalize but rather to provide a rich, contextualized understanding of some aspect of human experience through the intensive study of particular cases. Yet, in an environment where evidence for improving practice is held in high esteem, generalization in relation to knowledge claims merits careful attention by both qualitative and quantitative researchers. Issues relating to generalization are, however, often ignored or misrepresented by both groups of researchers. Three models of generalization, as proposed in a seminal article by Firestone, are discussed in this paper: classic sample-to-population (statistical) generalization, analytic generalization, and case-to-case transfer (transferability). Suggestions for enhancing the capacity for generalization in terms of all three models are offered. The suggestions cover such issues as planned replication, sampling strategies, systematic reviews, reflexivity and higher-order conceptualization, thick description, mixed methods research, and the RE-AIM framework within pragmatic trials.

© 2010 Elsevier Ltd. All rights reserved.

What is already known about the topic?

- The topic of generalization is less often discussed by qualitative than by quantitative researchers, who consider the ability to generalize a key quality criterion.
- Many leaders in qualitative research have begun to note the importance of addressing generalization, to ensure that insights from qualitative inquiry are recognized as important sources of evidence for practice.

What this paper adds

- Generalization can be clarified by recognizing that there are three different models of generalization, each of which has relevance to nursing research and evidence-based practice: the classic statistical generalization model, analytic generalization, and the case-to-case transfer model (transferability).
- Both quantitative and qualitative researchers uphold certain myths about adherence to the three models of generalization, and these myths hinder the likelihood that real opportunities for generalization will be pursued.
- Many strategies can be adopted by both qualitative and quantitative nurse researchers to enrich the readiness of their studies for reasonable extrapolation.

Generalization is an act of reasoning that involves drawing broad conclusions from particular instances—that is, making an inference about the unobserved based on the observed. In nursing and other applied health research, generalizations are critical to the interest of applying the findings to people, situations, and times other than those in

* Corresponding author at: Humanalysis, Inc., 75 Clinton Street, Saratoga Springs, NY 12866, United States. Tel.: +1 518 587 3994; fax: +1 518 583 7907.

E-mail address: dpolit@rocketmail.com (D.F. Polit).

a study. Without generalization, there would be no evidence-based practice: research evidence can be used only if it has some relevance to settings and people outside of the contexts studied.

Although many articles and books have discussed the issue of generalizability, few have considered strategies for addressing it in nursing research. The purpose of this paper is to discuss three different models of generalization, to identify "myths" about the degree to which these models are adhered to in qualitative and quantitative research, and to offer suggestions for enhancing the capacity for generalization in nursing research.

1. Introduction

In quantitative research, generalizability is considered a major criterion for evaluating the quality of a study (Kerlinger and Lee, 2000; Polit and Beck, 2008). Within the classic validity framework of Cook and Campbell (e.g., Shadish et al., 2002), external validity—the degree to which inferences from a study can be generalized—has been a valued standard for decades. Yet, generalizability is a thorny, complex, and illusive issue even in studies that are considered to yield high-quality evidence (Kerlinger and Lee, 2000; Shadish et al., 2002).

In qualitative studies, the issue of generalization is even more complicated, and more controversial. Qualitative researchers seldom worry explicitly about the issue of generalizability. The goal of most qualitative studies is to provide a rich, contextualized understanding of human experience through the intensive study of particular cases. Qualitative researchers do not all agree, however, about the importance or attainability of generalizability. Some challenge the possibility of generalizability in any type of research, be it qualitative or quantitative. In this view, generalization requires extrapolation that can never be fully justified because findings are always embedded within a context. According to this way of thinking, knowledge is idiographic, to be found in the particulars (Guba, 1978; Erlandson et al., 1993). On the other hand, some qualitative researchers believe that in-depth qualitative research is especially well suited for revealing higher-level concepts and theories that are not unique to a particular participant or setting (Glaser, 2002; Misco, 2007). In this view, the rich, highly detailed, and potentially insightful nature of qualitative findings make them especially suitable for extrapolation.

In the current evidence-based practice environment, the issue of the applicability of research findings beyond the particular people who took part in a study has gained importance for qualitative researchers. Groleau et al. (2009), in discussing generalizability in a recent article in *Qualitative Health Research*, argued that an important goal of qualitative studies is to shape the opinion of decision-makers whose actions affect people's health and well-being. Thorne (2008) echoed similar sentiments about the need to adopt a practical perspective: "…the moral mandate of a practice discipline requires usable general knowledge…(Qualitative) researchers in this field are obliged to consider their findings 'as if' they might indeed be applied in practice" (p. 227). Ayres et al. (2003)

observed that, "Just as with statistical analysis, the end product of qualitative analysis is a generalization, regardless of the language used to describe it" (p. 881).

2. Models of generalization

Firestone (1993) developed a typology depicting three models of generalizability that provides a useful framework for considering generalizations in quantitative and qualitative studies. The first model is extrapolating from a sample to a population (statistical generalization), the classical model underpinning most quantitative studies. The second model is analytic generalization, a model that has relevance in both qualitative and quantitative research. The third model is case-to-case translation, which is more often called transferability. The latter two models have been described as mechanisms for dealing with the apparent paradox of qualitative research—its focus on the particular and its simultaneous interest in the general and abstract (Schwandt, 1997).

2.1. Statistical generalization

In the familiar model of generalization—what Lincoln and Guba (1985) referred to as nomothetic generalization—quantitative researchers begin by identifying the population to which they wish to generalize their results. The population is the totality of elements or people that have common, defined characteristics, and about whom the study results are relevant. Researchers proceed to select participants from that population, with the goal of selecting a sample that is *representative* of the population. The best strategy for achieving a representative sample is to use probability (random) methods of sampling, which give every member of the population an equal chance to be included in the study with a determinable probability of selection. Standard tests of statistical inference are based on the assumption that random sampling from the target population has occurred (Polit, 2010).

Like most models, this generalizability model is an *ideal*—a goal to be achieved, rather than an accurate depiction of what transpires in real-world research. Yet the myth that this model is adhered to in quantitative scientific inquiry in the human sciences perseveres.

One flaw stems from the starting point: most quantitative researchers begin with only a vague notion of a target population. They are more likely to have an explicit accessible population, that is, a group to which they have access and from which participants are sampled. Even accessible populations, which are linked to hypothetical target populations in a diffuse and often unarticulated way, frequently are ill-defined in research reports. In many cases, the population may be identified based on sample characteristics and relevant eligibility criteria—that is, the real starting point is often the sample, not the population.

Random sampling is the vehicle through which the statistical model of generalization can be enacted. Even a casual perusal of journal articles in nursing and health care is sufficient to conclude that the vast majority of studies

with human beings do not involve random samples. Intervention studies, in particular, almost never rely on random samples (Shadish et al., 2002). In the rare study in which participants are sampled at random, cooperation is rarely perfect, which means that random *sampling* seldom results in random *samples*. Yet, the myth of random sampling in quantitative research persists. For example, Teddlie and Tashakkori (2009), in their recent textbook on mixed methods research, described quantitative sampling as "mostly probability" (p. 22). The random sampling myth is one in which virtually all researchers conspire when they apply standard statistical tests to analyze their data, in violation of the assumption of random sampling.

2.2. Analytic generalization

In analytic generalization, Firestone's second model of generalization, researchers strive to generalize from particulars to broader constructs or theory. Analytic generalization is most often linked with qualitative research, although it is implicitly embedded within theory-driven quantitative research as well.

In an idealized model of analytic generalization, qualitative researchers develop conceptualizations of processes and human experiences through in-depth scrutiny and higher-order abstraction. In the course of their analysis, qualitative researchers distinguish between information that is relevant to all (or many) study participants, in contrast to aspects of the experience that are unique to particular participants (Ayres et al., 2003). Generalizing to a theory or conceptualization is a matter of identifying evidence that supports that conceptualization (Firestone, 1993).

Analytic generalization in qualitative inquiry occurs most keenly at the point of analysis and interpretation. Through rigorous inductive analysis, together with the use of confirmatory strategies that address the credibility of the conclusions, qualitative researchers can arrive at insightful, inductive generalizations regarding the phenomenon under study. As noted by Thorne et al. (2009), "When articulated in a manner that is authentic and credible to the reader, (findings) can reflect valid descriptions of sufficient richness and depth that their products warrant a degree of generalizability *in relation to a field of understanding*" (p. 1385, emphasis added).

As is true for statistical generalizability, the analytic generalization model is an ideal that is not always realized. Thorne and Darbyshire (2005), in their provocative and clever paper designed to inspire improvements in qualitative health research, noted a number of tendencies of qualitative researchers that undermine analytic generalization. Examples of the problematic patterns they identified based on years of experience as qualitative researchers included premature closure ("stopping at the 'aha'"), enthusiasm for artificial coherence ("fitness addiction"), and stopping when it is convenient rather than when saturation is attained ("the wet diaper," p. 1108). They specifically noted that problematic qualitative health reports present "overgeneralizations that spill out from the conclusions," (p. 1107) and then become fodder for the critics of qualitative research.

2.3. Transferability

The third model of generalizability proposed by Firestone (1993) is what he called *case-to-case translation*. Case-to-case transfer, which involves the use of findings from an inquiry to a completely different group of people or setting, is more widely referred to as *transferability* (Lincoln and Guba, 1985), but has also been called *reader generalizability* (Misco, 2007).

Transferability is most often discussed as a collaborative enterprise. The researcher's job is to provide detailed descriptions that allow readers to make inferences about extrapolating the findings to other settings. The main work of transferability, however, is done by readers and consumers of research. Their job is to evaluate the extent to which the findings apply to new situations. It is the readers and users of research who "transfer" the results.

Transferability has close connections to concepts developed by research methodologist Donald Campbell (1986), who suggested an approach to generalizability called the proximal similarity model. Campbell himself thought that *proximal similarity* was a more suitable term than *external validity*—a term he himself had coined—for considering how research might be extrapolated. Within the proximal similarity model, researchers and consumers envision which contexts are more or less like the one in the study. His model involves conceptualizing a *gradient of similarity* for times, people, settings, and contexts, from most closely similar to least similar. Proximal similarity supports transferability to those people, settings, sociopolitical contexts, and times that are most like (i.e., most proximally similar to) those in the focal study. A similar idea was suggested by Lincoln and Guba (1985), who used the term *fittingness* to refer to the degree of congruence or similarity between two contexts.

Although transferability is a concept that has been used as a quality criterion primarily in qualitative research, the proximal similarity model brings to light the salience of transferability in quantitative research as well. Let us assume, for example, that a random sample of women from Ohio participated in a health survey. Findings about the correlation between (for example) poverty and smoking in that sample could be generalized to women throughout the state, according to the statistical model of generalization. The proximal similarity (transferability) model is appropriate for considering whether the findings could be extrapolated to women in other midwestern states, to women in Australia, and to women in rural China—or to men in Ohio.

In discussing strategies to support transferability, most writers discuss the need for *thick description* (Geertz, 1973; Lincoln and Guba, 1985). Thick description refers to rich, thorough descriptive information about the research setting, study participants, and observed transactions and processes. Readers can make good judgments about the proximal similarity of study contexts and their own environments only if researchers provide high-quality descriptive information. As Firestone (1990) noted, thick description is not restricted to prose, as the name implies, but involves all forms of critical information (including

demographic information) that helps readers to understand the study's context and participants.

The transferability model, like the previous two models of generalizability, represents an idealized goal for researchers. In reality, the kind of description that supports transferability is often not as "thick" as readers need for making informed judgments about proximal similarity. Journal articles, which have tight page constraints, are not "friends" of thick description. Journal page limits alone are not to blame, however. In a recent analysis of over 1000 journal articles published in eight top-tier nursing journals in 2005 and 2006, nearly 20% of the papers did not report the sex distribution of the sample, two-thirds did not describe the racial or ethnic distribution, and nearly half failed to provide information about participants' average age. Qualitative reports lacked such information more often than quantitative reports (Polit and Beck, 2009). The absence of such basic information about participants suggests that thick description may be another myth that merits attention.

Of course, *thick description* means more than reporting sample characteristics. In qualitative research in particular, thick description requires rich description of the study context and of the phenomenon itself. Yet, Thorne and Darbyshire (2005), as well as Sandelowski (1997), noted that some qualitative reports are woefully "thin" in terms of describing the phenomenon and in providing supporting data for the researchers' interpretations.

3. Strategies to enhance generalized inferences

Although it would be impossible to catalogue and discuss the many strategies that could be used to enhance generalization in nursing studies, we offer a few suggestions. Some are appropriate for individual researchers, and others are more global strategies.

3.1. Replication in sampling

Replication plays a role in all three models of generalization. In discussing analytic generalization, Firestone argued that "When conditions vary, successful replication contributes to generalizability. Similar results under different conditions illustrate the robustness of the finding" (p. 17).

Replication is an important principle in making sampling decisions. In qualitative research, various purposive sampling strategies that involve deliberate replication can be used to promote both analytic generalization and transferability. For example, critical case sampling, which involves selecting important replicates that illuminate critical aspects of a phenomenon (Patton, 2002), can contribute to the researchers' ability to crystallize a conceptualization. Replication with deviant cases (i.e., the most unusual or extreme informants) can help to refine or revise a conceptualization, but can also help to understand extreme conditions under which the conceptualization holds. Replication within a range of cases that vary on attributes likely to affect conceptualization (maximum variation sampling) can also help to strengthen generalization.

Saturation of important themes and categories is the sampling principle used to enhance the likelihood that analytic generalization can occur, and saturation clearly encompasses the concept of replication. As Morse et al. (2002) pointed out, there must be sufficient, and redundant, information in qualitative research to account for all aspects of a phenomenon. Strategic replication in qualitative research can also enhance transferability.

In quantitative research, replication of participants, in the form of adding to sample size, can enhance generalization, as well as statistical power. Even when samples are not drawn at random, the more replicates there are, the greater the likelihood that unusual cases will cancel each other out, which in turn can contribute to the sample's representativeness—although the famous example of the sample of *Literary Digest* readers who led to the prediction than Alfred Landon would defeat Franklin Roosevelt in the 1936 presidential election reminds us that large samples can also harbor bias.

Small convenience samples of participants who are not selected for any theoretical reasons are all too common in quantitative studies, and yet it is precisely this type of design that poses the most severe threats to the conventional model of generalizability. Indeed, it could be argued that quantitative researchers would do better at achieving representative samples for the statistical generalizability model if they had a more purposive approach—that is, if they explicitly added replicates to correspond more closely to population parameters. Quota sampling, for example, is a semi-purposive sampling strategy that is far superior to convenience sampling because it seeks to ensure sufficient replicates within key strata of the population. Another purposive replication strategy for enhancing representativeness is multi-site sampling. Shadish et al. (2002) also argued for more purposive sampling, noting that deliberate heterogeneous sampling on presumptively important dimensions is an important strategy for generalization.

3.2. Replication of studies

At a broader level, there needs to be greater encouragement for the planned replication of studies (Fahs et al., 2003), which enhances the potential for generalizability in all three models. If concepts, relationships, patterns, and successful interventions can be confirmed in multiple contexts, varied times, and with different types of people, confidence in their validity and applicability will be strengthened. Indeed, the more diverse the contexts and populations, the greater will be the ability to sort out "irrelevancies" from general truths (Shadish et al., 2002). Yet, deliberate replication is often not seen as valuable, and is sometimes actively discouraged for graduate students. Knowledge does not come simply by testing a new theory, using a new instrument, or inventing a new construct (or, worse, giving an inventive label to an old construct). Knowledge grows through confirmation. Many theses and dissertations would likely have a bigger impact on nursing practice if they were replications that yielded systematic, confirmatory evidence—or if they revealed restrictions on generalized conclusions. Of course, in both qualitative and

quantitative research, intentional replication only makes sense when there are strong, thoughtful studies that yield evidence worth repeating.

3.3. Integration of evidence

Perhaps the most important development for enhancing generalizations in health care research comes from the recent methodologic and conceptual advances in integrating evidence from multiple studies. Systematic integration in the form of meta-analysis and metasynthesis has emerged as a cornerstone of the evidence-based practice movement. Such integration relies on replications.

Meta-analysis involves the statistical integration of results from multiple quantitative studies addressing the same research question. Meta-analyses are a good way to address, if not redeem, the deficiencies of individual studies with regard to the classic generalizability model. By combining results from participants of different backgrounds, settings, time periods, and circumstances, greater clarity about generalized inferences and limits to those inferences can be achieved.

Metasynthesis involves interpretive translations produced from the integration of findings about a phenomenon from multiple qualitative studies. Analytic generalization is particularly well exemplified in high-quality metasyntheses because higher-order abstractions are often developed, and the conceptual power of the formulation about the phenomenon of interest is enhanced. Thorne (2009), in discussing metasynthesis, noted that "findings from distinct studies in a field can be rigorously integrated into stronger and more generalizable knowledge claims." Schofield (1990) recognized the potential of integration for enhancing generalizability in qualitative research two decades ago, well before modern-day methods for metasynthesis evolved.

Although systematic integrations of qualitative and quantitative findings bode well for the analytic generalization and statistical models of generalization, it remains to be seen whether integrative summaries can play a role in the transferability model. Integrations tend to strip out information about study contexts, and may offer little guidance about the extent to which the generalizations developed in the integration can be transferred. Typically, integrative reviews include only limited descriptions for assessing proximal similarity. Future methodological work on integration might tackle this important problem of transferability. Moreover, the potential for integration to contribute to generalizable knowledge cannot be realized if the integration is poorly conceived or executed without recognition of the complexity of the task.

3.4. Thinking conceptually and reflexively

All models of generalization can benefit from researchers spending more time reflecting on their concepts and data, rather than focusing disproportionate attention on methods. Nearly five decades ago, the philosopher of social science Arthur Kaplan (1964) worried about the pervasive characteristic he called *methodolatry*, the "overemphasis on what methodology can achieve" (p. 24). The primacy of method over substance is a concern echoed by several qualitative researchers (e.g., Sandelowski, 1997), and is relevant to the discussion of generalizability.

A generalization inherently involves abstraction of general concepts from particular observations. Generalization is something in which all human being engage. Stake (1978), for example, used the term *naturalistic generalization* to refer to generalizations as a product of human experiences in which people identify "the similarities of objects and issues" (p. 6). Eisner (1998), who wrote at length about generalization and qualitative inquiry, made a similar point. He observed that, "Human beings have the spectacular capacity to go beyond the information given, to fill in gaps, to generate interpretations, to extrapolate, and to make inferences in order to construe meanings. Through this process, knowledge is accumulated, perception is refined, and meaning deepened" (p. 211). Researchers can enhance the likelihood that such generalization and knowledge accumulation will happen by being more reflexive, by thinking conceptually, and by using strategies to enhance the potential for all three types of generalization.

Quantitative researchers are perhaps more guilty than qualitative researchers of not paying attention to conceptual matters in an ongoing way. In many quantitative studies, researchers relegate the bulk of their conceptual energies to the early "conceptual phase" (Polit and Beck, 2008) of a study. Once the intellectual and creative work of formulating a problem, theoretical context, and study design has been completed, the implementation of the research plan can sometimes be rather mechanical. Yet, in the midst of data collection, thoughtful reflection about the setting, the participants, and the data themselves could foster insights that would contribute to generalized understandings. In many quantitative studies, there is also room for improvement during the "conceptual phase" for developing a strong theoretical or conceptual basis, to enhance analytic generalization.

To do high-quality work, qualitative researchers must be reflexive and conceptual throughout their project. Their emergent efforts to ask good questions of the right people (or to observe the right behaviors or events) force ongoing decisions that are, in theory at least, driven by the conceptual demands of the study, and it is these efforts that contribute to analytic generalization. Quantitative researchers likely would benefit from more thoroughly understanding qualitative methods, and applying some to their own research.

Conceptualization is clearly an aspect of analytic generalization, but is relevant in considering transferability and the proximal similarity model as well. Because researchers are familiar with only the "sending contexts" of their study and not the "receiving contexts" of potential users (Lincoln and Guba, 1985), some argue that the researcher's responsibility is solely to provide thorough description of the sending contexts so that transferability becomes an option for readers. The proximal similarity model suggests, however, that researchers can go a bit further. In developing their thick descriptions, researchers can think conceptually rather than simply descriptively about their contexts. That is, they can develop (and

communicate) a theoretical perspective about essential contextual features that might make their findings transferable so that readers can make theoretically-informed judgments about which contexts are most proximally similar. The goal is not so much to have a formal theory about contexts and gradients of similarity, as to have a framework that is abstract and conceptual for deciding on the types of descriptive information to share. This recommendation applies to quantitative as well as qualitative researchers.

Relatedly, Greenwood and Levin (2005) suggested reframing generalization as a process involving reflective action by users of research. In this active process of reflection, people decide for themselves whether or not previous findings make sense in a new context. In Greenwood and Levin's two-step model for generalizing knowledge to new settings, a potential user first strives to conceptualize the context under which the findings were created. Then in the second step, he or she must understand the contextual conditions of the new setting. Reflection is involved as the person considers the consequences of applying the findings to the new context.

3.5. "Know thy Data"

Qualitative researchers are expected to be *immersed* in their data. Immersion in and strong reflection about one's data can promote effective generalization, particularly for the analytic generalization model but also for the other generalization models.

The process of "making meaning" and developing powerful analytic generalizations in qualitative studies relies on the researcher's thorough understanding of and engagement with the data. Ayres et al. (2003) provided an excellent discussion of how analytic generalization (they used the term *generalizability*) can be strengthened through intensive within-case and across-case analysis. They noted that such immersion does not always occur, and that some researchers "fail to go beyond the production of a list of themes or key categories" (p. 881). Without being thoroughly absorbed with one's own data and with the details of the study context, researchers may also fall short in providing the thick descriptions upon which transferability depends.

Quantitative researchers could benefit from greater immersion in their data as well. Indeed, quantitative researchers are often disconnected from their data in ways that can undermine their capacity for insightful interpretation and generalization. For example, in large studies quantitative researchers often do not collect their own data but rather hire research assistants to do this. Many do not analyze their own data either, relying instead on statistical consultants. Having technical assistance and staff support is essential in large-scale studies, but this should not prevent lead researchers from getting close to their data.

Even when working with large data sets, quantitative researchers can get to know their participants and study contexts better by looking at data horizontally (within cases), rather then simply vertically (across cases for specific variables). Scrutiny of the complete data record for

extreme cases, typical cases, exemplary cases, and people from small subgroups can often illuminate "what is going on" in a dataset in a way that computing a statistic cannot. This process has been referred to as the "qualitizing" of quantitative data (Sandelowski, 2000). Sandelowski (2001) has also noted that the "quantitizing" of qualitative data can serve to confirm conclusions and to generate meaning, which can promote analytic generalization.

Even in the absence of qualitizing, quantitative researchers can enhance their generalizability claims by developing a better familiarity with who is in their samples—and who is not. Given that the validity of classic generalizations depends on seldom-achieved random sampling, Kerlinger and Lee (2000) noted that the model can work reasonably well "if we are careful about studying our data to detect substantial sample idiosyncrasy" (p. 286). They advised researchers to check their sample for easily verified expectations. For example, if half the population is known to be female, then the researcher can check to see if approximately half the sample is female.

3.6. Thick description

There is broad agreement that description needs to be sufficiently detailed to permit transferability. Yet, even Lincoln and Guba (1986) acknowledged that "it is by no means clear how 'thick' a thick description needs to be" (p. 77).

Decisions about degrees of "thickness" will depend on the particulars of the research, but a general recommendation is for researchers to consciously consider the consequences of their "thickness" decision for the applicability of their evidence. Qualitative and quantitative researchers need to do a better job at providing basic information about their participants, contexts, and time-frames. Readers should know when data were collected, what type of community was involved, and who the participants were, in terms of their age, gender, race or ethnicity, and any clinical or social characteristics that might affect an assessment of proximal similarity.

Descriptive information is sometimes hidden from readers in an effort to protect confidentiality, but sometimes withholding information serves little purpose. For example, researchers are more likely to say that their study was done "in a large American city" than to say that it was done in, say, Boston. In most studies, especially quantitative ones, no one's confidentiality would be breached by communicating the specific locale—and yet, naming the city could offer valuable information for potential users' judgments about proximal similarity. In any event, many readers easily can infer that the research setting was Boston based on the author's institutional affiliation—suggesting perhaps yet another myth within research circles. The withholding of information about specific locales does not appear to be mandated within ethical guidelines, with the possible exception of case studies (American Psychological Association, 2010). Thus, researchers should consider sharing precise information about the context of their studies, unless there are prohibitions about doing so from institutional partners.

3.7. Mixed methods research

Mixed methods research, which involves the collection, analysis, and integration of qualitative and quantitative data within a study or coordinated series of studies, appears to hold promise for generalizability. Larger and more representative samples in the quantitative strand of mixed methods studies can promote confidence in generalizability in the classic sense. Well-grounded meta-inferences (Teddlie and Tashakkori, 2009) based on rich, complementary data sources can enhance analytic generalization. And rich and diverse descriptive information from two types of data source can promote an understanding of proximal similarities and hence transferability. Interest in mixed methods research is growing rapidly, and exciting developments are also occurring with regard to mixed methods integration (e.g., Flemming, 2010; Plueye et al., 2009). It remains to be seen, however, whether mixed methods research will live up to the promise of enhancing generalization potential. To a very large extent, this will depend on strategic and judicious blending of data to arrive at knowledge that merits generalization.

3.8. Pragmatic trials and the RE-AIM framework

In intervention research, the tension between internal validity (the ability to infer a causal link between an intervention and measured outcomes) and external validity (generalizing effects beyond the study sample) has long been a source of consternation (Shadish et al., 2002). The traditional solution has been to sacrifice external validity to internal validity by designing tightly controlled randomized controlled trials (RCTs) with stringent exclusion criteria that can render the results difficult to apply in the real world.

Another solution to the validity conflict is a phased approach in which researchers first conduct controlled *efficacy RCTs* (wherein internal validity is first established), followed by *effectiveness studies* (wherein the generalizability of effects is tested under more realistic circumstances). Although this approach has appeal, it has seldom been used in nursing—effectiveness studies are costly and rare.

Recently, there have been calls for undertaking "pragmatic" clinical trials that strive to achieve a balance between internal and external validity in a single trial (Glasgow et al., 2005; Borglin and Richards, 2010). Relatedly, efforts to improve the generalizability of evidence from intervention studies have given rise to a framework called RE-AIM (Glasgow, 2006). The RE-AIM framework involves a scrutiny of five aspects of an intervention trial: its Reach, Efficacy, Adoption, Implementation, and Maintenance. Two elements of the framework, Reach and Adoption, directly relate to generalizations. *Reach* concerns the extent to which study participants have characteristics that reflect those of that population. *Adoption* concerns the number and representativeness of settings in which the intervention is adopted. Use of the RE-AIM framework is likely to enhance generalization, especially within the statistical generalization and transferability models. The RE-AIM framework is beginning to be adopted in nursing intervention studies (Whittemore et al., 2009).

4. Discussion

Generalizability or applicability is an issue of great importance in all forms of health and social research, and this is particularly true in the current environment in which *evidence* is held in high esteem. Qualitative and quantitative studies have developed their own special ways of dealing with generalization, none of them with perfect success. Arguably though, there are fewer "myths" relating to the analytic generalization and transferability models than to the statistical model of generalizability that has been cherished as a criterion of excellence in quantitative research. The statistical generalizability model is almost *never* fully realized, even though the research community usually acts as though it is. It is therefore somewhat ironic that critics of constructivist approaches often cite "the generalizability problem" as a critical factor for not giving qualitative research its due (Sandelowski, 1997).

Like all models in the behavioral and social sciences, the three models of generalization discussed in this paper are ideals, not representations of reality. And, in fact, leading thinkers and methodologists from both the postpositivist and constructivist schools have long recognized that generalizations can never be made with certainty. For example, a prominent measurement expert (Cronbach, 1975) and an influential advocate for naturalistic approaches (Guba, 1978) both asserted that any generalization represents a *working hypothesis*. Cronbach noted that, "When we give proper weight to local conditions, any generalization is a working hypothesis, not a conclusion" (p. 125). Guba concurred, writing that "in the spirit of naturalistic inquiry (the researcher) should regard each possible generalization only as a working hypothesis, to be tested again in the next encounter and again in the encounter after that" (p. 70).

Kerlinger and Lee (2000) advanced a similar position in discussing generalizability not as an absolute but as something that exists on a continuum. In discussing the classic model of generalizability, they noted that the usual question of whether the results of a study can be generalized to other people or settings should perhaps be replaced with a question of relativity: "*How much* can we generalize the results of the study?" (p. 474, emphasis in original).

Despite the difficulty in both qualitative and quantitative research of achieving the ideals embodied in the three models of generalization, they offer a good frame of reference for planning and conducting research. The standard statistical model of generalization may not be relevant for qualitative researchers, but all three models of generalization are germane in quantitative research—although this is seldom recognized. To develop evidence that is useful to practitioners, nurse researchers should strive to meet the generalization ideals embodied in the models, to compensate for lapses from it, and to identify those lapses so that the worth of study evidence can be more accurately assessed. They must also, of course, strive to develop evidence with high validity and integrity, which is a precondition to any generalizability goal.

With evidence-based practice gaining increasing acceptance, the eyes of the entire nursing community are on

the evidence that nurse researchers produce—both in terms of its validity/trustworthiness and its potential for application in real-world settings. Rather than disdaining the possibility of generalizability (some qualitative researchers) or unfairly assailing the limitations of qualitative research to yield general truths (some quantitative researchers), researchers with roots in all paradigms can take steps to enrich the readiness of their studies for "reasonable extrapolation" (Patton, 2002, p. 489).

Of course, clinicians will always need to be thoughtful about using "generalizable" evidence, because generalizations are never universal. As Lincoln and Guba (1985) noted, "The trouble with generalizations is that they don't apply to particulars" (p. 110). Donmoyer (1990) also cautioned against directly generalizing from research findings to specific individuals in specific circumstances. Evidence with high potential for generalizability represents a good starting starting point—a working hypothesis that must be evaluated within a context of clinical expertise and patient preferences.

Conflict of interest: None declared.

Funding: None.

Ethical approval: Not applicable – no human subjects.

References

American Psychogical Association, 2010. Publication Manual of the American Psychological Association, 6th ed. Author, Washington, DC.

Ayres, L., Kavanagh, K., Knafl, K., 2003. Within-case and across-case approaches to qualitative data analysis. Qualitative Health Research 13, 871–883.

Borglin, G., Richards, D., 2010. Bias in experimental research: strategies to improve the quality and explanatory power of nursing science. International Journal of Nursing Studies 47, 123–128.

Campbell, D.T., 1986. Relabeling internal and external validity for the applied social sciences. In: Trochim, W. (Ed.), Advances in Quasi-Experimental Design and Analysis. Jossey-Bass, San Francisco, pp. 67–77.

Cronbach, L., 1975. Beyond the two disciplines of scientific psychology. American Psychologist 30, 116–127.

Donmoyer, R., 1990. Generalizability and the single-case study. In: Eisner, E.W., Peshkin, A. (Eds.), Qualitative Inquiry in Education: The Continuing Debate. Teachers College Press, New York, pp. 175–200.

Eisner, E.W., 1998. The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice. Prentice Hall, Upper Saddle River, NJ.

Erlandson, D., Harris, E., Skipper, B., Allen, S., 1993. Doing Naturalistic Inquiry: A Guide to Methods. Sage, Thousand Oaks, CA.

Fahs, P., Morgan, L., Kalman, M., 2003. A call for replication. Journal of Nursing Scholarship 35, 67–72.

Firestone, W.A., 1990. Accommodation: toward a paradigm-praxis dialectic. In: Guba, E. (Ed.), The Paradigm Dialog. Sage, Newbury Park, CA, pp. 105–124.

Firestone, W.A., 1993. Alternative arguments for generalizing from data as applied to qualitative research. Educational Researcher 22, 16–23.

Flemming, K., 2010. Synthesis of quantitative and qualitative research: an example using critical interpretive synthesis. Journal of Advanced Nursing 66 (1), 201–217.

Geertz, C., 1973. Thick description: toward an interpretive theory of culture. In: Geertz, C. (Ed.), The Interpretation of Cultures. Basic Books, New York.

Glaser, B.G., 2002. Conceptualisation: on theory and theorizing using grounded theory. In: Bron, A., Schemmenn, M. (Eds.), Social Science Theories and Adult Education Research. Lit Verlag, Munster, pp. 313–335.

Glasgow, R.E., 2006. RE-AIMing research for application: ways to improve evidence for family medicine. Journal of the American Board of Family Medicine 19, 11–19.

Glasgow, R.E., Magid, D., Beck, A., Ritzwoller, D., Estabrooks, P., 2005. Practical clinical trials for translating research to practice: design and measurement recommendations. Medical Care 43, 551–557.

Greenwood, D.J., Levin, M., 2005. Reform of the social sciences and of universities through action research. In: Denzin, N.K., Lincoln, Y.S. (Eds.), The Sage Handbook of Qualitative Research. Sage Publications, Thousand Oaks, CA, pp. 43–64.

Groleau, D., Zelkowitz, P., Cabral, I., 2009. Enhancing generalizability: moving from an intimate to a political voice. Qualitative Health Research 19, 416–426.

Guba, E., 1978. Toward a Methodology of Naturalistic Inquiry in Educational Evaluations. University of California, Center for the Study of Evaluation, Los Angeles.

Kaplan, A., 1964. The Conduct of Inquiry. Chandler, Scranton, PA.

Kerlinger, F.N., Lee, H.B., 2000. Foundations of Behavioral Research, 4th ed. Harcourt College Publishers, Fort Worth, TX.

Lincoln, Y., Guba, E., 1985. Naturalistic Inquiry. Sage, Beverly Hills, CA.

Lincoln, Y., Guba, E., 1986. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. In: Willimas, D. (Ed.), Naturalistic Evaluation. Jossey Bass, San Francisco, pp. 73–84.

Misco, T., 2007. The frustrations of reader generalizability and grounded theory: alternative considerations for transferability. Journal of Research Practice 3, 1–11.

Morse, J.M., Barrett, M., Mayan, M., Olson, K., Spiers, J., 2002. Verification strategies for establishing reliability and validity in qualitative research. International Journal of Qualitative Methods 1 (2) Article 2. Retrieved January 7, 2010 from http://www.ualberta.ca/ijqm.

Patton, M.Q., 2002. Qualitative Evaluation and Research Methods, 3rd ed. Sage, Thousand Oaks, CA.

Plueye, P., Gagnon, M., Griffiths, F., Johnson_Lafleur, J., 2009. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. International Journal of Nursing Studies 46, 529–546.

Polit, D.F., 2010. Statistics and Data Analysis for Nursing Research, 2nd ed. Pearson Education, Upper Saddle River, NJ.

Polit, D.F., Beck, C.T., 2008. Nursing Research: Generating and Assessing Evidence for Nursing Practice, 8th ed. Lippincott Williams & Wilkins, Philadelphia, PA.

Polit, D.F., Beck, C.T., 2009. International differences in nursing research, 2005–2006. Journal of Nursing Scholarship 41, 44–53.

Sandelowski, M., 1997. "To be of use": enhancing the utility of qualitative research. Nursing Outlook 45, 125–132.

Sandelowski, M., 2000. Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. Research in Nursing & Health 23, 246–255.

Sandelowski, M., 2001. Real qualitative researchers do not count: the use of numbers in qualitative research. Research in Nursing & Health 24, 230–240.

Schofield, J.W., 1990. Increasing the generalizability of qualitative research. In: Eisner, E.W., Peshkin, A. (Eds.), Qualitative Inquiry in Education: The Continuing Debate. Teachers College Press, New York, pp. 201–232.

Schwandt, T.A., 1997. Qualitative Inquiry: A Dictionary of Terms. Sage, Thousand Oaks, CA.

Shadish, W., Cook, T., Campbell, D., 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin, Boston.

Stake, R.E., 1978. The case study method in social inquiry. Educational Researcher 7, 5–8.

Teddlie, C., Tashakkori, A., 2009. Foundations of Mixed Methods Research. Sage, Thousand Oaks, CA.

Thorne, S., 2008. Interpretive Description. Left Coast Press, Walnut Creek, CA.

Thorne, S., 2009. The role of qualitative research within an evidence-based context: can metasynthesis be the answer. International Journal of Nursing Studies 46, 569–575.

Thorne, S., Darbyshire, P., 2005. Land mines in the field: a modest proposal for improving the craft of qualitative health research. Qualitative Health Research 15, 1105–1113.

Thorne, S., Armstrong, E., Harris, S., Hislop, T., Kim-Sung, C., Oglov, V., et al., 2009. Patient real-time and 12-month retrospective perceptions of difficult communications in the cancer diagnostic period. Qualitative Health Research 19, 1383–1394.

Whittemore, R., Melkus, G., Wagner, G., Dziura, J., Northrup, V., Grey, M., 2009. Translating the diabetes prevention program to primary care. Nursing Research 58, 2–12.