
Qualitative and Quantitative Evidence in Health: The Critics' View

6

Kadriye Ercikan and Wolff-Michael Roth

6.1 Introduction

Much of what medical researchers conclude in their studies is misleading, exaggerated, or flat-out wrong.

Freedman (2010)

This is the beginning sentence of an article in the *Atlantic* magazine entitled “Lies, Damned Lies, and Medical Sciences.” The entire article is about Dr. John Ioannidis, whose highly cited studies have been challenging medical research with similarly provoking claims about the accuracy of medical research results reported in reputable journals. For example, Ioannidis (2005) boldly states that most current medical research findings are false. The objects of his criticism are not limited to small-scale observation-based research studies with small sample sizes but also include what might be thought of as gold standards of research: randomized control trials (RCTs). In addition to the criticisms about accu-

racy of claims from RCTs, the external validity of RCTs has been questioned. Six main issues that may affect the generalizability of knowledge claims from RCTs have been identified (Rothwell 2005): (a) setting of the trial, (b) selection of patients, (c) characteristics of randomized patients, (d) differences between the trial protocol and routine practice, (e) outcome measures and follow-up, and (f) adverse effects of treatment. Underlying this chapter is our conviction that it is more important for health researchers to worry about the *quality* of research evidence than about whether the research is of the quantitative, qualitative, or mixed-method type. Independent of whether the research is qualitative or quantitative, we are as critical of research that overgeneralizes as we are of research that fails to offer generalizations beyond the actual case(s) studied.

The question of what constitutes “good” research evidence is at the heart of this chapter. Evidence, as is the term is typically used in evidence-based practice, refers to “an observation, fact, or organized body of information offered to support or justify inferences or beliefs in the demonstration of some proposition or matter at issue” (Upshur 2001, p. 7). Such research inferences typically involve making *generalizations* to populations and contexts beyond the specific samples used in the research. The question about generalizability of research findings—also referred to as external validity—is not simple, cut

K. Ercikan, Ph.D. (✉)

Cross-Cultural Assessment & Research Methods in Education, Faculty of Education, University of British Columbia, Vancouver, BC, Canada
e-mail: Kadriye.ercikan@ubc.ca

W.-M. Roth, Ph.D.

Applied Cognitive Science, University of Victoria, Victoria, BC, Canada

and dry, because the results' experimental studies and epidemiological studies *may not pertain* to the individual precisely because of their generalizing nature; and the results of some qualitative research approaches pertain to *every* individual, despite the extreme reluctance of many qualitatively working researchers to seek generalization of research results (Ercikan and Roth 2014; Roth 2009b). The ultimate questions about health research are these: "to whom do the knowledge claims from research apply?" and "if others cannot use the findings of my research, if my results do not generalize to other settings, what good is it to try reporting them?"

Evidence-based practice informed by empirical research is highly emphasized in many fields: at the time of this writing (June 2013), a simple search for the topic "evidence-based" in the ISI Thomson database yielded 48,230 articles. The first five of these articles are from health-related fields (occupational health, health education, nursing). The term "evidence-based," however, often has been taken as synonymous with the results of experimental research and large-scale statistical studies, whereas the results from observational and qualitative studies have been thought of as providing anecdotal evidence only. Yet the question "what constitutes evidence?" is much more complex than the association between evidence and statistical/experimental research.

On the one hand, quantitative research has well-established guidelines for determining what counts as evidence and which research findings can be generalized. These generalizations typically are tied to the research design (such as in RCTs) and representativeness of the sample relative to the target population of interest. The question remains whether the same evidence supports decisions for different groups and individuals. To answer these questions, we need to consider the extent to which health research conducted in one setting (a) can be used to inform other settings and (b) findings generalize from one sample to the target population and, thereby, apply to another subpopulation and other individuals. In a policy context that places great value on evidence-based research, experimental research and research using high-power statistics tend to be

privileged as having the capacity to support generalizations that can contribute to decision-making in policy and practice (e.g., Slavin 2008; Song and Herman 2010). Group-level evidence—such as whether an intervention is effective based on an experimental design—likely is not sufficient to make decisions about effectiveness of the intervention for individuals or for subgroups such as males or females, age groups, or individuals from different ethnic and racial backgrounds. This is not only so because of the statistical nature of experimental design but also because almost all experiments are based on the logic of interindividual differences and covariations rather than on a logic of within-individual differences and causations (Borsboom et al. 2003). Moreover, a recent Bayesian analysis of 855 published studies in experimental psychology showed that in 70 % of the cases with *p*-values between 0.01 and 0.05, "the evidence is only *anecdotal*" (Wetzels et al. 2011, p. 291, emphasis added).

In qualitative research, on the other hand, claims are often restricted to the settings, to the subjects, and to the context of the research without efforts to derive generalizable claims from research. Thus, for example, one qualitative study concerned with pediatric oncology education reports positive impact and yet calls for "further evidence" that "truly analyses the effectiveness and impact of education on paediatric oncology practice" (McInally et al. 2012, p. 498). That is, there is a contradiction within the article that both claims to report on the impact of oncology education *and* calls for studies that *truly* analyze the impact.¹ This does not have to be this way, because there are ways of going about qualitative research such that the results are invariant across all members of a population; that is, they pertain to every individual.

¹This study was picked at random from the sample of articles that resulted from a search in the ISI Thomson Web of Science database with the search parameters "evidence based," "qualitative research," and "health." We do not "pick" on this particular article but use it as an example of a general tendency among qualitative researchers to refuse seeking the *general* rather than sticking with the particular.

Moreover, in some forms of research, such as the Bayesian approach, statistical and qualitative information is explicitly combined. Thus, one study concerned with identifying factors that mediate adherence to medication regimens in HIV situations synthesized a body of quantitative and qualitative studies by generating qualitative themes emerging from the former and by quantizing the information from the latter and then used Bayesian data augmentation methods to summarize all studies (Crandell et al. 2011). In this chapter, we articulate the structure and discuss limitations of different forms of generalizations across the spectrum of quantitative and qualitative research and argue for a set of criteria for evaluating research generalization and evidence.

6.2 Generalization in Quantitative and Qualitative Research

Similar to social science research, *generalization* in health research is a critical concept where the specific is expanded to the general (“generalized”) and the general is reduced to the specific (“particularized”) in the creation of knowledge to inform policy and practice (Ercikan and Roth 2009; Roe 2012). In both qualitative and quantitative research, generalization typically focuses narrowly on representativeness of the sample on individuals of the population the generalizations target. However, in addition to the target population, many facets of research determine generalization, including: time and context of research, attributes focused on, and methods utilized. The degree to which knowledge claims from research can inform practitioners who deal with individuals and policymakers who deal with groups depends on the degree to which such contexts, methods, and attributes are applicable to the target generalization situation. In this section we discuss three main forms of generalizations: *analytic*, *probabilistic*, and *essentialist*. We highlight their limitations in view of how they can be considered as evidence to inform policy and practice.

6.2.1 Analytic Generalization

Analytic generalization relies on the design of the research to allow making causal claims, for example, about the effectiveness of a health intervention (Shadish et al. 2002). The primary logic in this design is this: instances where a cause operates have to lead to significantly different observations than those instances where the cause is disabled. The design requires randomly assigning participants to control and experimental groups in the hope of achieving equivalence of these groups with respect to all moderating and mediating variables and an identical implementation of the intervention to the experimental group. The experimental and control groups are not expected to be representative samples of any particular target population. Instead, random equivalence of these two groups is central to the experimental design and is intended to rule out any potential alternative explanations of differences between the two groups. The causal claims from analytic generalization are closely tied to the degree to which the experimental design truly implements the theoretical relations between causes and effects. The statistical support for the effectiveness of the treatment is determined by comparing the difference between the mean outcome scores of control and experimental groups to the standard error of the mean differences. A statistically significant difference in the hypothesized direction between control and experimental groups provides evidence to support a causal claim about the effectiveness of the treatment.

Causal claims in analytic generalization are evaluated based on two key criteria. The first criterion is whether there is a systematic difference between experimental and control groups that can be supported by statistical evidence. The second criterion is the degree to which a true experiment has been conducted so that the change in experimental group outcomes can be attributed to the specific operating cause deriving from the treatment and explained by theory. The causal claims in analytic generalization is based on the logic of between-subjects rather than within-subjects variation (Borsboom et al. 2003) and can be supported only at the overall group level. In

other words, treatment may have been effective “on the average” but the causal claim may not apply to some individuals or some subgroups. Figure 6.1 presents distributions of outcome scores for experimental and control groups from a hypothetical experiment. As the overlapping area in Fig. 6.1 shows, a considerable number of individuals in the control group may have higher scores than individuals in the experimental group. Even though individuals from the experimental group are more likely to be on the higher end of the outcome score scale and those from the control group are more likely to be at the lower end of the scale, we cannot tell how the change in scores varied for different individuals or subgroups and whether the change was uniformly in the same direction. The degree of change and the direction of change for individuals in the experimental group cannot be determined by comparing score distributions with the control group.

6.2.2 Probabilistic (Sample-to-Population) Generalization

Probabilistic generalization—also known as statistical or sample-to-population generalization—relies on representativeness of a sample of a target population. It is used to describe population characteristics and does not include causal claims (Eisenhart 2009; Yin 2008). Researchers and consumers of research judge knowledge claims by the degree to which samples of subjects, outcomes, and contexts used in research are representative of the populations to which the research is intended to generalize (Ercikan 2009; Firestone 1993). Two broad types of probabilistic generalizations are common. One type of generalization claim is with respect to relationships between variables. An example of such research includes an investigation of the relationship between anxiety and suicide attempts based on a nationally representative data of US adult population from the National Epidemiologic Survey on Alcohol and Related Conditions Wave 2 (NESARC II) (Nepon et al. 2010). This research demonstrated that of all those who made a suicide attempt, over 70 % had at least one anxiety

disorder. In this research, statistics is used to estimate the probability that a systematic relation between each disorder and suicide attempt exists beyond chance level. The second type of research generalization is related to relative frequency for demographic or other groups of interest. For example in the Nepon et al. (2010) study, these generalizations include the proportion of individuals identified with anxiety disorders or suicide attempts by gender groups. In both of these probabilistic generalizations, generalization claims are derived from observations from the sample. The criteria by which the generalization is judged—i.e., the validity of claims about the correlation between anxiety disorders and suicide attempts or gender differences in anxiety disorders—center on one of the same criteria used for judging analytic generalization, that is, whether there is statistical evidence of a systematic pattern in the data. Even though probabilistic generalizations may include group comparison, such as comparing gender or ethnic groups, these generalizations do not require a specific research design such as random equivalence of groups or standardized implementation of an intervention. Instead, the representativeness of the samples of the target populations is the second key criterion used for probabilistic generalizations.

Within-group heterogeneity limits the meaningfulness of *causal* claims in analytic generalization for subgroups or individuals and leads to similar limitations in probabilistic generalization.

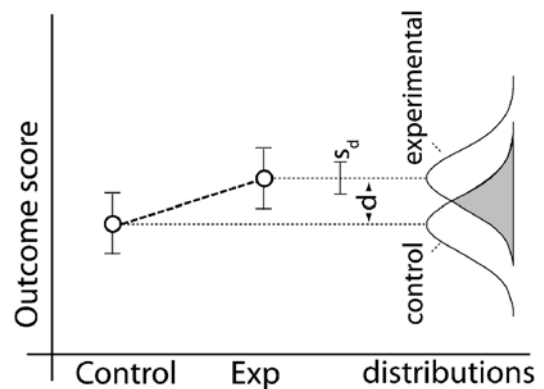


Fig. 6.1 Hypothetical distribution of outcome scores for experimental and control groups

When there is great diversity within the target population, such diversity will be reflected in the nationally representative sample the research is based on. It has been argued that cultures differ in fundamental aspects such as reasoning styles, conceptions of the self, the importance of choice, notions of fairness, and even visual perception (Henrich et al. 2010). Research claims regarding such psychological constructs at national levels will have limited generalizability to different cultural groups. It is easy to see how prevalence statistics, for example, for anxiety disorders, may vary for subgroups, such as people with illnesses, gender, or age groups. In fact, one study found that in cancer patients the risk of psychiatric distress was nearly twice that of the general population (Hinz et al. 2010). Therefore, a claim about prevalence of illnesses—a probabilistic generalization—has limitations in its accuracy and meaningfulness for different subgroups. Population heterogeneity may lead to similar problems in generalizations of correlational relationships. Based on their research, Nepon et al. (2010) conclude that panic disorders are associated with suicide attempts. These researchers established evidence that individuals who are diagnosed with panic disorder are two and a half times more likely to attempt suicide than those who are not diagnosed with this disorder. Research has demonstrated a great degree of variation in the prevalence of suicide attempts across cultures—participating countries were the United States, Canada, Puerto Rico, France, West Germany, Lebanon, Taiwan, Korea, and New Zealand—ranging from 0.72 in Lebanon to 5.93 in Puerto Rico (Weissman et al. 1999). Such diversity limits the applicability of single prevalence statistics across cultures. It is also expected to affect the degree to which correlation between suicide attempts and other variables. In summary, therefore, when population heterogeneity is present, probabilistic generalization focusing on describing population characteristics can lead to knowledge claims that involve statistical concepts—e.g., mean, frequency, mean differences, or correlations—that may not apply to subgroups and may have limited value for guiding policy and practice.

6.2.3 Essentialist Generalization

Essentialist generalization systematically interrogates “the particular case by constituting it as a ‘particular instance of the possible’ [...] to extract general or invariant properties that can be uncovered only by such interrogation” (Bourdieu 1992, p. 233). Because *every* particular is as good as any other, such research therefore identifies invariants of the phenomenon that hold across *all* particulars related to the phenomenon of interest. Therefore, essentialist generalization involves identifying aspects of the phenomenon that applies to *all* individuals in the population. The identification of invariants, and therefore the construction of a generalization, is possible by focusing on the *process* by means of which a phenomenon manifests itself rather than on the manifestations themselves (as this would be reported in phenomenographic studies). Classical examples of essentialist generalization derive from studies within the dialectical tradition, which seeks to understand the diversity of social life and phenomena based on cultural-historical and evolutionary precursors. An example of such research is the Russian psychologist L. S. Vygotsky’s (1971) generation of a *general theory* of the psychology of art. He was interested not in the psychology of any particular art form but of art in general: “*I talk about all art and do not verify my conclusions on music, painting, etc.*” (Vygotsky 1927/1997, p. 319, original emphasis). Thus, he took as his “very special task to find *the precise factual boundaries* of a general principle in practice and the *degree* to which it can be applied to different species of the given genus” (p. 319).

In his analysis, Vygotsky begins with one fable and, having articulated some general principles that make up the basis of all art forms, uses *one* short story and *one* tragedy as a test-bed of his findings (Fig. 6.2). Just as stated in the above quotation, Vygotsky took the particular case of the fable and isolated with his analysis *affective contradiction* and *catharsis* as the essential processes in/of any art form. This required him to “abstract from the concrete characteristics of the fable as such, a specific genre, and [to] concentrate

the forces upon the *essence* of the aesthetic reaction” (Vygotsky 1927/1997, p. 319, our emphasis). That is, he ascended in a way the tree and located properties that are typical not only in the specific art form but in all forms of art that descended from the same essential aesthetic reaction (Fig. 6.2). This essence is true for and generalizes to all art forms. For this reason, we refer to such generalization as *essentialist*. Commenting on the approach, the author of the introduction to *The Psychology of Art*—who himself has used the essentialist method to trace the origin of human emotion to the first forms of life, such as, single-cellular organisms—notes that:

[...] the significance and function of a poem about sorrow is not at all to transmit the author’s sadness to the reader [...] but that it changes this sorrow in such a manner as to reveal something new and pertinent to man on a higher level of truth.
Leontiev in Vygotsky (1971, p. vii)

To arrive at the essence of art, the analysis focused on the “aesthetic reaction,” that is, on “*the processes in their essence*” (Vygotsky 1927/1997, p. 319). In concrete contexts, which for the psychology of arts constitute the different

art forms, the *essential* processes bring forth phenomena that *appear* different (e.g., fable, pottery, blues music, Fig. 6.2), that seem to constitute difference, when in fact the processes are the same. The author concludes that this method is similar to the classical experiment: the “meaning” of the result “is broader than its field of observation” (p. 319), though the principle “*never manifested itself* in pure form, but always with its ‘coefficient of specification’” (p. 319).

Pertaining to health research, one can find this essentialist generalization in a philosophical tradition of E. Husserl, such as the analysis of organ transplants, beginning with the experience of: (a) receiving a new heart (Nancy 2000) or a new liver (Varela 2001), (b) long-term chronic fatigue syndrome (Roth 2009a, 2014), (c) taking psychoactive drugs (Roth 2011), or (d) suffering in general (Roth 2011). In these instances, the analyses do not strive to communicate the singular experiences of these authors/patients but are designed to reveal fundamental processes and phenomena that underlie the experience of an organ transplant *generally*, including “the lived body and its exploration, the unalienable alterity of our lives,

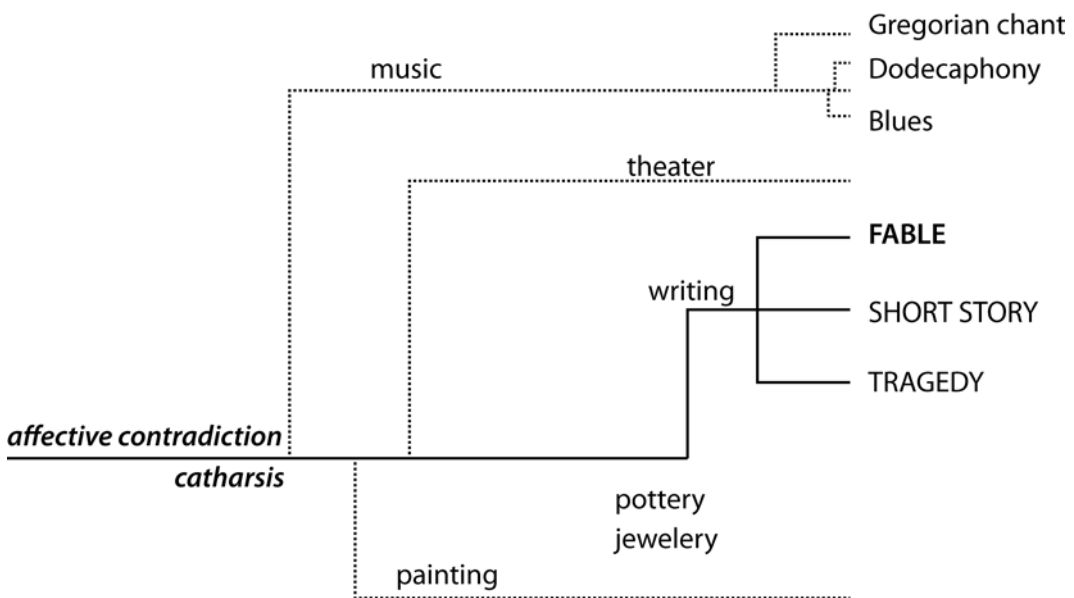


Fig. 6.2 Vygotsky’s essentialist generalization derived a general law for the psychology of art on the basis of one case (here the *fable*); the law was tested in a small number of other cases (here *short story* and *tragedy*)

the key ground of temporality, body-technologies and ethics" (p. 271). The purpose of this type of phenomenological study is to arrive at descriptions of experience that make it possible to allow collaboration with those "hard sciences" that investigate true causes (i.e., at the individual level) rather than impute causes that only describe relations at the group level. For this reason, the "*disciplined first-person* accounts should be an integral element of the validation of a neurobiological proposal and not merely coincidental or heuristic information" (Varela 1996, p. 344, original emphasis, underline added). That is, there is an explicit rapprochement of "hard" experimental work and "soft" (but essentialist) qualitative research; and there are calls for the explicit coordination of research combining essentialism with forms of research in the natural and experimentally working social sciences (Bourdieu 1992; Roth and Jornet 2013).

Recent studies have shown that first- and second-person phenomenological methods in physical and emotional health research may be used to arrive at generalizations of phenomena that may be identified in the study of one individual but that are observable in *every* individual case (e.g., Roth 2012; Vygotsky 1927/1997). One study concerned with predicting the occurrence of epileptic seizures exhibited the possibility to, "establish correlations between precise 'first person' descriptions of the subjective experience of a given cognitive process [...] and 'third person' measures of the corresponding neuro-electric activity" (Petitmengin et al. 2006, p. 299). Thus, the correct identification of signs of seizures can be used for alternative, non-medication-based therapies that actually prevent or interrupt a seizure. Other studies from a neuro-phenomenological perspective exhibit the viability of connecting the qualitative descriptions of experience with hard evidence from EEG and fMRI studies of physiological and psychiatric phenomena (e.g., Micoulaud-Franchi et al. 2012).

Research conducted in this vein identifies in the singular case a particular instance *of the possible*. This possible constitutes the general. In this way, the general is as concrete as the particular. The following genetic analogy further speci-

fies this relation. The genes of the parents constitute the possible with respect to their offspring. Even though all children may look different, they constitute a particular instance of the possible. If we tried to identify the general in an inductive way, by analyzing the identical features of all children, we may not be able to identify any commonality (Il'enkov 1982). Rather, essentialist generalizations are found "through *analysis* of at least one typical case rather than through *abstraction* of those identical features that all possible cases have in common" (p. 170).

One approach to qualitative research that pursues an agenda of identifying general processes that lead to situated particulars is *ethnomethodology* (Garfinkel 1967). As its name suggests, the approach investigates the methods by means of which ordinary people constitute, in concert with others, the structured everyday world of our experience—e.g., practices underlying sex change, "bad" clinic records, or psychiatric outpatient clinic selection. Rather than concerning itself with the panoply of a type of social phenomenon, such as queues that exist in a multitude of ways—e.g., at a movie ticket counter, supermarket cash register, highway ramp, or bus stop—ethnomethodology is concerned with the methods people use and make visible to each other in lining up, recognizing beginnings and ends of queues, identifying problems with queuing, and so on. These methods underlying queuing transcend any particular lineup and, therefore, constitute the general. The identification of these methods does not require special research methods because every person competently lining up implicitly practices them. In this way, ethnomethodology is a radical alternate method that is incommensurable with all other standard—qualitative or quantitative—researches (Garfinkel 2007). It is a radical alternate because it provides answers to the question: what more is there to social practices than what all qualitative or quantitative formal analytical research has established? (Roth 2009c). The tremendous opportunities and promises arising from this approach for health research have been recognized and outlined but have yet to be realized in research practice (e.g., Dowling 2007).

6.3 Quantitative Research: The Critic's View I

6.3.1 Assumptions in Generalization

Research generalization involves making some assumptions in order for knowledge claims based on specific research to apply to individuals and contexts not involved in the research. Three key assumptions in generalizations from both social science and health research may be identified: (a) *uniformity-of-nature*, (b) *continuity*, and (c) *ceteris paribus* assumptions (Roe 2012). The *uniformity-of-nature* assumption posits that all people are similar in their properties and behaviors and are hence exchangeable. The *continuity assumption* refers to constancy of individuals' characteristics and behavior and to the fact that these do not change over time. This makes it possible to examine them any time. The *ceteris paribus assumption* refers to constancy of other factors and hence to assume the possibility of their influence on characteristics and behaviors being investigated.

Uniformity-of-Nature Assumption Most inferential research assumes invariance of constructs, behaviors, and processes among individuals. Excluding clear constraints—e.g., adults, people living in rural or urban areas, etc.—that are targeted by the research, this assumption leads to the situation that all people are possible candidates for the research and its claims. As long as the individuals meet the broad categories, heterogeneity within these groups is overlooked or neglected. The *uniformity-of-nature assumption* has its roots in the natural sciences where it may hold reasonably well for physical characteristics. However, this assumption cannot be expected to hold for human beings who are affected by, and *react to*, the physical and social environments they live in.

Continuity Assumption Another key assumption in most inferential research is that human characteristics are invariant over time. This leads to

researchers investigating human characteristics without taking time into account. Time can include seasons of the year, time of the day, periods such as decades, etc. Very little research focuses on change over time as the targeted construct (Roth and Jornet 2013). This assumption does not match reality where clear differences are observed between generations and eras.

Ceteris Paribus Assumption Researchers typically manipulate a limited number of factors in their research and generalizations are made assuming that “all other things are equal.” However, we have to ask how reasonable it is to assume that all other things can possibly be equal? Researchers frame their generalization either by arguing that the results are invariant under other conditions or may caution that the results may be different under different conditions. To the degree that individuals vary in different settings, over time, and the research effects vary under different conditions, violation of these assumptions will lead to inaccurate generalizations.

In the next section we provide examples of research where these three assumptions are implicitly made and where there is evidence of violation of these assumptions. In addition, we provide examples of psychological research that lacks explicit identification of target generalizations.

6.3.2 Violation of Generalization Assumptions and Lack of Explicit Identification of Target Generalizations

Even when generalization to people is implicit, typical research does not even identify and describe to which population the results are intended to generalize. In fact, in typical research, researchers start with the sample and representativeness of the sample is either not recognized at all or recognized when the results are discussed after the fact (Roe 2012). Lack of reference to a population exists even in research that includes inferential statistics, which implies claims

targeted to a population. For example, an abovementioned study conducted with adults in 2004 and 2005 concludes that “[a]nxiety disorders, especially panic disorder and PTSD, are independently associated with suicide attempts. Clinicians need to assess suicidal behavior among patients presenting with anxiety problems” (p. 791). This research claim makes no reference to which adult population this conclusion applies to and under what conditions. Without an explicit statement of what the results are intended to generalize to, however, it is impossible to know to whom the findings may apply and what the limits of the generalization are. Replicability or verifiability of such studies is also limited if the researchers do not know from which population samples should be drawn.

Time is another facet of research that is typically not considered in generalizations (though it is explicitly theorized in cultural-historical approaches to psychology). Time is expected to be a factor in populations, contexts, settings, or interventions. Individuals’ psychological structures in different decades and periods may be different. For example, IQ measures and personality type tend to be taken as invariants. The sociocultural context of the research undergoes continuous change and, thereby, may influence how people perceive and react to things. Interventions that may have been effective in the 1960s—e.g., women struggling with self-esteem issues—may be irrelevant and ineffective for women today. For example, one study describes gender differences in self-esteem (SE) based on empirical research without any reference to its temporal context:

Three experiments explored the idea that men’s and women’s SE arise, in part, from different sources. It was hypothesized that SE is related to successfully measuring up to culturally mandated, gender-appropriate norms—separation and independence for men and connection and interdependence for women. Results from Study 1 suggested that men’s SE can be linked to a individuation process in which one’s personal distinguishing achievements are emphasized. Results from Study 2 suggested that women’s SE can be linked to a process in which connections and attachments to important others are emphasized. Study 3 demonstrated that failing to perform well on gender-

appropriate tasks engendered a defensive, compensatory reaction, but only in subjects with high SE.

Josephs et al. (1992, p. 391)

But can we really make the assumption that self-esteem issues do not change over time within the same population? In fact, in current research practices not contextualizing interpretation of research findings in temporal contexts is the norm, not the exception.

6.3.3 Implications for Determining Evidence

Evidence-based practice involves using research results to inform decisions affecting groups or individuals. Violation of the three assumptions above leads to inaccurate inferences from research for evidence-based practice. In essence, such a process involves generalizing from the “universal” to the “particular.” How can we determine if a treatment that was effective on the average will be effective for the individual? Such an inference ignores individual differences and changes over time. These inferences ignore heterogeneity in populations as well as effects of time on research generalizations leading to inaccurate inferences. Improving generalizations will require identifying the extent to which the assumptions hold. Only then researchers can determine whether the research provides evidence for the individual(s), time, and contexts, for evidence-based practice.

6.4 Qualitative Health Research: The Critic’s View II

In this section, we take a critical look at current qualitative health research methods with respect to the criteria that have been articulated for determining its quality and with respect to the methods of arriving at true generalizations through qualitative research that we articulate in Sect. 6.2. The purpose of the critique is to prepare qualitative health research to make a step toward taking their responsibility in the research communica-

tion and translation process, clearly identifying and articulating invariants across settings to be expected in their results just as we expect this to be the case for good quantitative research. We briefly analyze a randomly selected article.

In qualitative research, generalization tends to be thought of differently from that used in experimental research and large-scale statistical studies and sometimes authors are explicit about non-generalizability of their findings. For example, there are researchers claiming to do “phenomenological studies” focusing on the experience of one individual or a few individuals. This is in evident contradiction to the phenomenological method reviewed above, which is designed to arrive at understandings that are valid *for every* human being—e.g., in the work of Husserl, Heidegger, or Merleau-Ponty—right up to the present day, for example, in the phenomenological analyses of the experiences of organ transplants conducted by the affected individuals themselves (Nancy 2000; Varela 2001). If, however, research results were not generalizable to some extent, then these would not transfer (could not be transported) to and thereby inform a new context. If research results can inform settings other than those in which the research has been conducted, then it behooves the authors to articulate which other settings might benefit and the limitations that occur in the transfer of claims between settings. To distinguish the nature of quality criteria that differentiate qualitative from quantitative studies, a set of parallel criteria has been proposed (Lincoln and Guba 1985). Subsequently, Guba and Lincoln (1989) presented a set of criteria that should be used for judging the quality of qualitative research.

In *naturalistic inquiry* (Lincoln and Guba 1985), the authors explicitly reject the idea of generalizability, arguing that it is a positivistic idea. They argue for a move to the question of *transferability*, which denotes the extent to which findings from one qualitatively studied setting can be transferred to another qualitatively studied setting. They rightly suggest that transferability cannot be based on the sending context alone but requires an understanding of the receiving con-

text. This, however, is nothing other than thinking about the user and usage as one criterion of generalizability (Ercikan and Roth 2014). Whereas it may be correct that investigators who know only the sending context cannot make generalizability (transferability) inferences, it is also correct that investigators interested in publishing their findings need to know just what in their findings is of interest to others generally, and to readers of the journal article more specifically. In this move, they have thereby done a first step in generalization by extending their site-specific findings to the sites of interest to their readers. Other researchers focus on receiving contexts and refer to external validity of qualitative research as *recognizability* (Konradsen et al. 2013). These researchers define recognizability as “[t]he degree to which individuals are able to recognize their own experience or the experiences of others in the findings of a qualitative study” (p. 70) and identify four categories of recognizability: full, partial recognition, recognition in others, and no recognition.

Meta-ethnographic studies (e.g., Noblit and Hare 1988) constitute one possible way in which qualitative health research can be compared and contrasted across ethnographic contexts. This then leads to syntheses of an ensemble of studies that overcomes a proliferation of apparently independent studies. Thus, for example, one recent systematic study reviewed the findings on smoking during pregnancy that derived from 26 studies (involving 640 pregnant women) reported in 29 papers (Flemming et al. 2013). If there had not existed a sufficient degree of generalizability, these studies could not have been synthesized. Things and phenomena become comparable only when they are categorized in the same way; and categorization inherently constitutes abstraction from situational particulars (Kant 1956). It therefore does not suffice when qualitative health researchers do not indicate in which way their findings and distinctions are transportable to and relevant in other settings, as in “I hope that some of these distinctions resonate with health care providers” (van Manen 1998, p. 10). That is, researchers often refuse to state in which way

their findings are relevant to contexts other than their own use of the adjective “phenomenological.” Thus, rather than seeking to identify *general principles* (Vygotsky 1927/1997), studies tend to limit their findings to the particular sample in their study, for example, “*participants in this study* are pointing to a different understanding of the relationship between personal experience and fear” or “participants on a white water raft trip experienced fear which helped to cement a sense of self” (Brymer and Schweitzer 2012, p. 484).

In contrast, the idea underlying the classical phenomenological approach is to identify, as Vygotsky had done, the general principles of underlying processes. These only manifest themselves in different form because of contextual particulars. This is very different from phenomenographic research, for example, which identifies, categorizes, and describes forms of experience. Thus, one health-related phenomenographic study was interested in describing and characterizing “what women with rheumatoid arthritis (RA) and juvenile idiopathic arthritis (JIA) perceive as important in considering the performance of daily occupations to perceive good health” (Ottvall-Hammar and Håkansson 2013, p. 82). Because of their approach, the authors are at pains to ascertain the representativeness of their sample to achieve “heterogeneity.” Nevertheless, they state that a limitation of their study was that they “did not involve all demographics that could have been of interest” (p. 90). Moreover, they explicitly address the question of the trustworthiness of their study and, therefore, expose its possibility to be untrustworthy. Finally, “transportability”—i.e., the qualitative researcher’s equivalence to generalizability (Lincoln and Guba 1985)—is not ascertained on the part of the researchers but transferred to the journal audience, who is asked to “decide whether the results can be transferred to the reader’s own contexts or not” (Ottvall-Hammar and Håkansson 2013, p. 84). The authors point out the challenging part of their study: how to handle their own preconceptions and perceptions.

6.5 Conclusions

Now, we often proceed as if what counts as evidence was evident because we trust a cultural routine, most often imposed and inculcated through schooling (the famous “methodology” courses taught at American universities). The fetishism of “evidence” will sometimes lead one to reject empirical works that do not accept as self-evident the very definition of “evidence.” Bourdieu (1992, p. 225)

In this section, we take a step back, articulating similarities and differences between research methods presented and critiqued in Sects. 6.3 and 6.4. We also situate this discussion in the context of research methods that explicitly combine quantitative and qualitative research, such as the Bayesian approach, which is concerned with evaluating the probabilities of scientific hypotheses, given prior evidence of quantitative and qualitative nature and using new data for updating the probability of these hypotheses. The Bayesian approach is useful as it offers a means of formalizing prior beliefs—the identification of which requires a qualitative approach—and combining the results with quantitative studies for the purpose of supporting decision-making, risk assessment, diagnosis and prognosis, or health technology assessment (Barbini et al. 2013).

In the introductory quotation to this section, Bourdieu notes that an effort to extract general or invariant properties from the particular “is too often lacking in the work of historians” (p. 233). He attributes this to the definition of the historian’s task and a less ambitious or pretentious, as well as less demanding, task thrust upon their discipline than that thrust upon, for example, sociologists. This may be just as true for researchers who define themselves as working within a qualitative tradition, as if *method* was the defining characteristic of what a scholar does. Rigid adherence to this or that method does not take us any further in understanding evidence and generalizability. Thus, “methodological indictments are too often no more than a disguised way of making a virtue out of necessity, of feigning to

dismiss, to ignore in an active way, what one is ignorant of in fact” (Bourdieu 1992, p. 226).

Synthesizing research from multiple studies is an important way of generalizing research findings across settings. Such studies inherently are concerned with comparing findings, which can be done only at a more abstract level where studies are comparable. Here, the Bayesian approach in particular provides tremendous opportunities because it allows the integration of qualitative and quantitative information. Most frequently, the qualitative information enters Bayesian analysis in the characterization of prior beliefs that enter the determination of probabilities prior to the research, probabilities that are then updated as a result of a research project (e.g., Roberts et al. 2002). Instead, as one study of the factors that influence nonadherence/adherence to HIV medication regimens shows, the results of quantitative and qualitative studies can enter syntheses at the same level and with equal weight (Crandell et al. 2011). The method allows the “borrowing of information across studies” and thereby “makes this method uniquely suited to the case where a variable is more heavily covered in qualitative or quantitative studies” (p. 667). There are other meta-analytic methods as well. But, in our view, those that integrate qualitative and quantitative studies are of particular value as they force researchers to push the boundaries of quantitative work to consider the nature of variance, on the one side, and push the boundaries of qualitative work to consider those aspects that will be invariant across settings.

To end, we encourage readers to heed a commentary made in reference to sociology to health research. Our research is:

[...] too serious and too difficult for us to allow ourselves to mistake scientific *rigidity*, which is the nemesis of intelligence and invention, for scientific *rigor*, and thus to deprive ourselves of this or that resource available in the full panoply of intellectual traditions of our discipline.

Bourdieu (1992, p. 227)

References

- Barbini, E., Manzi, P., & Barbini, P. (2013). Bayesian approach in medicine and health management. In A. J. Rodriguez-Morales (Ed.), *Current topics in public health* (pp. 18–35). doi:10.5772/52402. Retrieved from <http://www.intechopen.com/books/current-topics-in-public-health/bayesian-approach-in-medicine-and-health-management>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.
- Bourdieu, P. (1992). The practice of reflexive sociology. In P. Bourdieu & L. J. D. Wacquant (Eds.), *An invitation to reflexive sociology* (pp. 216–260). Chicago, IL: University of Chicago Press.
- Brymer, E., & Schweitzer, R. (2012). Extreme sports are good for your health: a phenomenological understanding of fear and anxiety in extreme sport. *Journal of Health Psychology*, *18*, 477–487.
- Crandell, J. L., Voils, C. I., Chang, Y., & Sandelowski, M. (2011). Bayesian data augmentation methods for the synthesis of qualitative and quantitative research findings. *Quality & Quantity*, *45*, 653–669.
- Dowling, M. (2007). Ethnomethodology: Time for a revisit? A discussion paper. *International Journal of Nursing Studies*, *44*, 826–833.
- Eisenhart, M. (2009). Generalization from qualitative inquiry. In K. Ercikan & W.-M. Roth (Eds.), *Generalizing from educational research: Beyond qualitative and quantitative polarization* (pp. 51–66). New York, NY: Routledge.
- Ercikan, K. (2009). Limitations in sample to population generalizing. In K. Ercikan & M.-W. Roth (Eds.), *Generalizing in educational research: Beyond qualitative and quantitative polarization* (pp. 211–235). New York, NY: Routledge.
- Ercikan, K., & Roth, W.-M. (2009). *Generalizing from educational research: Beyond qualitative and quantitative polarization*. New York, NY: Routledge.
- Ercikan, K., & Roth, W.M. (2014). Limits of generalizing in education research: Why Criteria for Research Generalization should include population heterogeneity and users of knowledge claims. *Teachers College Record*, *116*(6), 1–28.
- Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher*, *22*(4), 16–23.
- Flemming, K., Graham, H., Heirs, M., Fox, D., & Sowden, A. (2013). Smoking in pregnancy: A systematic review of qualitative research of women who commence pregnancy as smokers. *Journal of Advanced Nursing*, *69*, 1023–1036.
- Freedman, D. H. (2010). Lies, damned lies, and medical science. *The Atlantic*, *306*, 76–84.

- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Garfinkel, H. (2007). Lebenswelt origins of the sciences: Working out Durkheim's aphorism. Book Two: Workplace and documentary diversity of ethnomethodological studies of work and sciences by ethnomethodology's authors: What did we do? What did we learn? *Human Studies*, 30, 9–56.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. Beverly Hills, CA: Sage.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466, 29.
- Hinz, A., Krauss, O., Hauss, J. P., Höckel, M., Kortmann, R. D., Stolzenburg, J. U., & Schwarz, R. (2010). Anxiety and depression in cancer patients compared with the general population. *European Journal of Cancer Care*, 19, 522–529.
- Il'enkov, E. V. (1982). *The dialectics of the abstract and the concrete in Marx's Capital*. Moscow, USSR: Progress Publishers.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, e124.
- Josephs, R. A., Markus, H. R., & Tafarodi, R. W. (1992). Gender and self-esteem. *Journal of Personality and Social Psychology*, 63, 391–402.
- Kant, I. (1956). *Werke Band II: Kritik der reinen Vernunft* [Works vol. 2: Critique of pure reason]. Wiesbaden: Insel.
- Konradsen, H., Kirkevold, M., & Olson, K. (2013). Recognizability: A strategy for assessing external validity and for facilitating knowledge transfer in qualitative research. *Advances in Nursing Science*, 36, 66–76.
- Lincoln, Y. S., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- McInally, W., Masters, S., & Key, S. (2012). The impact of paediatric oncology education on clinical practice: A phenomenological study. *European Journal of Oncology Nursing*, 16, 498–504.
- Micoulaud-Franchi, J.-A., Fakra, E., Cermolacce, M., & Vion-Dury, J. (2012). Towards a new approach of neurophysiology in clinical psychiatry: Functional magnetic resonance imaging neurofeedback applied to emotional dysfunctions. *Neurophysiologie Clinique—Clinical Neurophysiology*, 42, 79–94.
- Nancy, J.-L. (2000). *L'intrus [The intruder]*. Paris, France: Galilée.
- Nepon, J., Belik, S. L., Bolton, J., & Sareen, J. (2010). The relationship between anxiety disorders and suicide attempts: Findings from the National Epidemiologic Survey on Alcohol and Related Condition. *Depression and Anxiety*, 27, 791–798.
- Noblit, G., & Hare, R. (1988). *Meta-ethnography: Synthesizing qualitative studies*. London: Sage.
- Ottenvall-Hammar, I., & Håkansson, C. (2013). The importance for daily occupations of perceiving good health: Perceptions among women with rheumatic diseases. *Scandinavian Journal of Occupational Therapy*, 20, 82–92.
- Petitmengin, C., Baulac, M., & Navarro, V. (2006). Seizure anticipation: Are neurophenomenological approaches able to detect preictal symptoms? *Epilepsy and Behavior*, 9, 298–306.
- Roberts, K., Dixon-Woods, M., Fitzpatrick, R., Abrams, K., & Jones, D. R. (2002). Factors affecting uptake of childhood immunisation: An example of Bayesian synthesis of qualitative and quantitative evidence. *Lancet*, 360, 1596–1599.
- Roe, R. (2012). *Using referents to improve generalization in psychological research*. Paper presented at the Limits of Generalizing symposium at the 30th International Congress of Psychology, Cape Town, South Africa.
- Roth, W.-M. (2009a). Living with chronic illness: On (medical) science and scientific literacy in everyday life. In W.-M. Roth (Ed.), *Science education from people for people: Taking a stand(point)* (pp. 146–171). New York: Routledge.
- Roth, W.-M. (2009b). Phenomenological and dialectical perspectives on the relation between the general and the particular. In K. Ericikan & W.-M. Roth (Eds.), *Generalization in educational research* (pp. 235–260). New York: Routledge.
- Roth, W.-M. (2009c). Specifying the ethnomethodological “what more?”. *Cultural Studies of Science Education*, 4, 1–12.
- Roth, W.-M. (2011). *Passibility: At the limits of the constructivist metaphor*. Dordrecht, The Netherlands: Springer.
- Roth, W.-M. (2012). *First person methods: Towards an empirical phenomenology of experience*. Rotterdam, The Netherlands: Sense Publishers.
- Roth, W.-M. (2014). Personal health – personalized science: a new driver for science education? *International Journal of Science Education*, 36, 1434–1456.
- Roth, W.-M., & Jornet, A. G. (2013). Situated cognition. *WIREs Cognitive Science*, 4, 463–478. doi:10.1002/wcs.1242.
- Rothwell, P. M. (2005). External validity of randomised controlled trials: To whom do the results of this trial apply? *Lancet*, 365, 82–93.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.
- Slavin, R. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluation. *Educational Researcher*, 37(1), 5–14.
- Song, M., & Herman, R. (2010). Critical issues and common pitfalls in designing and conducting impact studies in education: Lessons learned from the What Works clearinghouse (phase I). *Educational Evaluation and Policy Analysis*, 32, 351–371.
- Upshur, R. E. G. (2001). The status of qualitative research as evidence. In J. M. Morse, J. Swanson, & A. J. Kuzul (Eds.), *The nature of qualitative evidence* (pp. 5–26). Thousand Oaks, CA: Sage.

- Van Manen, M. (1998). Modalities of body experience in illness and health. *Qualitative Health Research*, 8, 7–24.
- Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, 3, 330–349.
- Varela, F. J. (2001). Intimate distances: Fragments for a phenomenology of organ transplantation. *Journal of Consciousness Studies*, 8, 259–271.
- Vygotsky, L. S. (1927/1997): The historical meaning of the crisis in psychology: A methodological investigation. In W. R. Rieber & J. Wollock (Eds.), *The collected work of L. S. Vygotsky vol. 6* (pp. 233–343). New York, NY: Kluwer Academic/Plenum Publishers.
- Vygotsky, L. S. (1971). *Psychology of art* (Scripta Technica, Transl.). Cambridge, MA: MIT Press.
- Weissman, M. M., Bland, R. C., Canino, G. J., Greenwald, S., Hwu, H. G., Joyce, P. R., & Yeh, E. K. (1999). Prevalence of suicide ideation and suicide attempts in nine countries. *Psychological Medicine*, 29, 9–17.
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science*, 6, 291–298.
- Yin, R. K. (2008). *Case study research: Design and methods* (Vol. 5). Thousand Oaks, CA: Sage.