Advances in Nursing Science
Vol. 26, No. 4, pp. 257–273

Research Methodology for the Internet
External Validity (Generalizability)

Sheryl LaCoursiere, RN, PhD

The burgeoning use of the Internet for nursing research calls for a need to examine characteristics of electronic populations, and how they can be conceptualized relative to their physically based counterparts. The 1963 work of Campbell and Stanley, now 40 years old, is revisited in relation to external validity, and its applicability to today's world of quantitative and qualitative research. Internal and external differences in Internet populations are outlined, and a new threat to external validity is proposed: the threat of Networked Effects, which is composed of 4 subfactors: Co-occurring Interferences to Testing, Self-selection Mechanisms, Electronic Group Environments, and Cultural Differences. These Internet threats must be well understood when formulating research methodology. **Key words:** *computer communication networks*, *Internet*, *nursing research*, *online*, *qualitative*, *quantitative*, *research design*, *research methodology*, *sampling*, *World Wide Web*

THE POSSIBILITIES for research on the Internet can be very compelling. By using time-honored research principles, the researcher can forge new paths of knowledge to help gain a new understanding of emerging and existing phenomena. However, principles of validity, and the ultimate ability to create inference, must be addressed. In looking at current Internet research,[1,2] it is apparent that work needs to be done to link contemporary topics to enduring principles of validity, particularly references to Campbell and Stanley's 1963 definitive work, *Experimental and quasi-experimental designs for research*,[3] the "bible of validity." This article will inform researchers about validity issues for investigations conducted on the Internet, and hopefully, assist in the identification of issues that researchers may need to consider when conducting a study on the Internet.

Although Campbell and Stanley's seminal text[3] has withstood the test of time and is frequently cited in non–Internet-related literature, it was written in a simpler time when networked communications were not the norm. Since the advent of the Internet, however, new issues in validity have emerged. Electronic systems, particularly the World Wide Web, e-mail, electronic newsgroups, and discussion boards, offer new avenues to access information and influence users' attitudes, values, and behaviors. However, these electronic systems present new and unique challenges for the research process, particularly in relation to the physical presence of the researcher.

To bring the concept of validity forward into a fresh understanding, the work of Campbell and Stanley[3] will be used as the backdrop in this article to examine validity issues in Internet research. A systematic analysis of the theory related to the notion of external validity (generalizability), and implications

From the Center for Excellence in Chronic Illness Care, Yale University School of Nursing, New Haven, Conn.

The author thanks Karyl Burns, RN, PhD, for her assistance with the preliminary manuscript, and Peggy Chinn, RN, PhD, FAAN, Henrietta Bernal, RN, PhD, Cheryl Beck, DNSc, CNM, FAAN, and Regina Cusson, RN, PhD, for their review and suggestions.

Corresponding author: Sheryl LaCoursiere, RN, PhD, Center for Excellence in Chronic Illness Care, Yale University School of Nursing, 100 Church St S, PO Box 9740, New Haven, CT 06536 (e-mail: sheryl.lacoursiere@yale.edu).

for Internet research methodology will be presented. Threats to generalizability will be explored, as well as other issues in both quantitative and qualitative research processes. As the fundamentals are explained, applicability to Internet situations will be noted, and examples described. By examining the principles of Campbell and Stanley and extending their work to Internet investigations, this article should help those who have questioned the inferences drawn from Internet research to not only understand but also implement strategies that support the validity of such investigations.

A REVIEW OF VALIDITY

What is validity? Derived from the Latin term *validitas*, meaning strength, validity is "a concept of logic characteristic of sound arguments," which "must be carefully distinguished from the truth." It is "syntactical, that is it is a property of the system rather than individual arguments."[4(pp804–805)] According to Cook and Campbell, it is "the best available approximation to the truth or falsity of propositions, including propositions about cause . . . at best, one can know what has not yet been ruled out as false."[5(p37)] Thus, validity can never be proven, only argued,[6] and all conclusions must be understood to be approximate or tentative.[5]

THE CONTRAST BETWEEN INTERNAL AND EXTERNAL VALIDITY

External validity is discriminated from internal validity by the nature of its focus. While internal validity looks *in* on factors impacting the estimation of truth, external validity looks *outward*, to assess the potential conclusions that may be drawn from the research and their application within a population. With internal validity, the *main effect* is of primary interest, whereas with external validity, the *interaction* with the exposure is of primary concern. This has implications for selecting a research design. To test statistically for an interaction, an analysis of variance (ANOVA) could be used, for example. The research design would need to provide data that meets the assumptions of ANOVA. For a discussion of statistics and appropriate tests, the reader is referred to statistical texts.

Campbell and Stanley distinguish internal validity as the "basic minimum without which any experiment is uninterpretable: Did in fact the experimental treatments make a difference in this specific experimental instance?"[3(p5)] In contrast, external validity "asks the question of *generalizability*: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?"[3(p5)] Further, they state the interaction between the two: "While internal validity is the *sine qua non*, and while the question of *external validity*, like the question of inductive inference, is never completely answerable, the selection of designs strong in both types of validity is obviously our ideal. Both types of criteria are obviously important, even though they are frequently at odds in that features increasing one may jeopardize the other."[3(p5)] In this article, the terms external validity and generalizability are used interchangeably.

THE NOTION OF GENERALIZABILITY: WHAT DOES IT MEAN ON THE INTERNET?

The root word of generalizability is *general*, derived from the Latin *genus*, or class. What does it mean to be "general"? Webster's defines it as "1. of, for, or from all; not local, special or specialized. 2. of or for a whole genus, kind, etc. 3. widespread (general unrest). 4. most common; usual. 5. not specific or precise. 6. highest in rank."[7(p257)] In addition, the word *generalizability* takes the definition one step further, alluding to its import in research: "1. to state in terms of a general law. 2. to infer or derive (a general law) from particular instances. 3. to formulate general principles. 4. to talk in generalities."[7(p257)] Thus, it can be seen that when

a researcher generalizes, the goal is to take specific information, and apply it to the widest possible circumstances.

The Internet is unlike any system that came before it. The nature of the Internet itself implies a network. This network requires us to examine our traditional beliefs about cause and effect. Because of its vast series of interconnections, the concept of linear paths takes on new connotations that must be considered by the researcher. Although in a measurable sense, the distance from point A to point B can be defined, the biggest change is the *multiple possibilities* for arriving in the same location. For instance, a person may arrive at a given Web site by way of a search engine, another site, a referral in an e-mail, or by previous knowledge.

The phenomenon of multiple possibilities presents new challenges in generalizability. In Internet research, defining a target population toward which the researcher wishes to generalize becomes increasingly complex. For instance, if the target population is persons with cardiovascular disease, Internet sites where persons with cardiovascular disease are present must first be found; this may include venues such as Web discussion boards, newsgroups, and e-mail discussion lists. Recruiting subjects from these sites raises new issues and a myriad of questions. These questions can be divided into 2 categories: (1) questions about internal differences *within* the Internet population and (2) questions about external differences, or the Internet population versus a "real world" population.

Internal differences

There are many questions relative to internal and external differences within Internet populations that pertain directly to generalizability. Although these questions merit depth on their own, they are presented here to illustrate the issues that might arise when exploring generalizability in Internet research.

Questions about internal differences *within* the Internet population relate to the characteristics of users in relation to their use of the Internet. To explore internal and external differences and their influence on generalizability, the researcher must first know the scope of the Internet population. For instance, Internet users in the United States represent a varying percentage of the total population. As shown in Table 1, older age groups are least represented. Depending on the nature of the study, this underrepresentation may impact the generalizability of research findings to persons over 65 years old. Persons in this group may need to be specifically sought out for participation.

The example of age is relatively simple. However, the distinctions between different

Table 1. Comparison of ages of US daily Internet users and the general population, in millions of persons

Age	US population (in millions)	US population who has been online (in millions)	US population who has been online (%)	Daily US Internet users (in millions)	US population using the Internet daily (%)
18–29	47	27	58	15	32
30–49	86	46	54	28	33
50–64	42	15	36	10	24
65+	35	4	11	3	9

Note: Data for Internet users were obtained from The Pew Internet and American Life Project[8(pp11,130)]; data for the general population were obtained from the US Census Bureau.[9]

types of Internet users may not be readily apparent to beginning researchers. Upon further examination, subtle and stark differences between users of various online modalities may be noted. For example, consider some of the numerous questions that pertain to internal distinctions in online modality use:

- Are there differences between persons who use discussion boards versus those who use e-mail discussion lists?
- What about people who use both—are they different than those who use e-mail only?
- Are people who use methods that involve synchronous or real-time interaction, such as chat, different from those who use asynchronous communication, where they are not interacting at the same time?
- Are they more sociable? Less depressed?

External differences

Questions about the Internet population versus the "real world" cannot really be called "external" differences because an Internet user inhabits both worlds. The Internet population is a subset of the real world population, but how large a subset is unknown. Figure 1 illustrates how, for example, Internet users are a subset of the population of a city, a country, and the world at large, and do not overlap. However, within each Internet population, there is overlap in the modalities used within the Internet itself. As the Internet becomes more ubiquitous, the Internet population may become increasingly more like the real world. Questions that may be asked, and assessed for their relevance as possible confounding variables, are as follows:

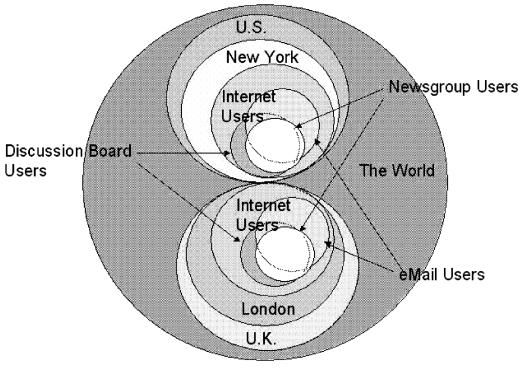

Fig 1. Sample populations of Internet users and their relationship to the world at large.

- How is the Internet population different demographically from the real world?
- How different are users in the United Kingdom compared to those in the United States?
- How different are users in London compared to those in New York?
- Is their socioeconomic status different?
- Does their status affect their access and/or knowledge of electronic resources?
- Is there a bias in favor of urban areas, where more electronic resources are available?
- From an international perspective, how are the persons from other countries who use the Internet culturally different from those of the United States?
- How are the persons from other countries who use the Internet culturally different from those in neighboring countries, for example China and Japan?
- How can the Internet population even be measured?
- Are the statistics from Internet measurement agencies reliable and valid?
- What is an appropriate comparison population? The persons who use the Internet resources, or their counterparts in the "real world?"
- Is there a difference in participants who complete one part of a study in one city, and another part in a different location? Where are they physically "located?"
- Is there an impact from family/friend/coworker presence in the room while the participant is completing the study?
- Does the relative anonymity of the Internet encourage more disclosure in this population, and how would that impact the research in question?
- How would it impact a quantitative as well as a qualitative study?

RESEARCH OBJECTIVES AND SAMPLING

In addition to factors specific to internal and external differences in the population, the researcher must determine if the objective for using the Internet in a particular study is for expediency or as a recruitment tool. For example, is the use of the Internet to avoid having to ask participants to come to a research center, or is it to collect data *unique to* the Internet, for example a web-based intervention? In either case, to determine generalizability for a specific Internet population, the researcher must make a judgment based on sound methodological principles, and best available information regarding threats to generalizability, as well as other issues to be considered, presented later in this article.

As can be seen, generalizability takes on new meaning when applied to an Internet population. One can see the difficulty of making inferences even *within* this population, without even regarding comparisons to physically present groups. From a statistical standpoint, it is known that as sample size increases, effect size and power increase and error decreases.[10] How will an *appropriate*, *representative* sample of the target population *with the desired characteristics* be obtained? If the researcher is attempting to obtain a sample of newsgroup users, how can it be determined that those who use newsgroups are similar to those who use newsgroups *and* e-mail? As can be seen, on the Internet, new *extraneous factors* that may affect the research may reach new dimensions.

The researcher must be very cognizant of the *objective* of the intended research, as well as the hypotheses to be tested, which will, of course, vary by discipline. For example, in marketing research, generalizability to a very large population may not be a desired goal. For instance, a medical device manufacturer makes a product specifically tailored for a certain group of users, for instance a blood sugar testing device for patients with diabetes with poor eyesight. In this case, the goal of research on the device would be to concentrate the assessment on those with visual deficits, and its impact on their health management. Whether or not the device has use or appeal to those with good eyesight would not be as important as its efficacy in the intended population. Thus, in this case the population would not be all diabetics, or those in need of blood glucose meters; it would be patients with diabetes with poor eyesight. If this research were conducted on the Internet, the population would be smaller, as it would be diabetic *Internet users* with poor eyesight. Thus, it would be more targeted, yet less generalizable. Carefully describing the population is necessary to allow for appropriate generalizability.

THREATS TO GENERALIZABILITY

Threats to external validity can be grouped into 5 categories: (1) the interaction of testing and X, (2) the interaction of selection and X, (3) other interactions with X, (4) multiple treatment interference, and (5) reactive arrangements.[3] In addition, a sixth factor is proposed: (6) threats related to extraneous networked effects. The first 3 of these threats are directly related to the 8 factors that are threats to *internal* validity: testing, selection, history, maturation, instrumentation, regression, mortality, and interactions (for example, selection and maturation). The next section of this article will review each threat to external validity, including specific application to Internet research.

The interaction of testing and X

Testing, a threat to internal validity, threatens generalizability when it *interacts with X*.*

*In experimental notation, an exposure, treatment, or intervention is designated by "X," and an observation such as the administration of a test, an "O." Threats pertaining to internal validity directly influence the *main effects* of the experiment, or the "O" scores, and are controlled by research designs. On the other hand, threats to external validity are considered *interaction effects*, that is they "involve X and some other variable.[3(p16)] This nomenclature is important when attempting to classify factors in the research process.

Testing is defined as "the effects of taking a test upon the scores of a second testing."[3(p5)] In research, pretests and posttests are used to measure the effect of an exposure. In some experiments, the content of the tests themselves may play a pivotal role in external validity. This is especially true in settings where pretests with unusual content may affect an individual's attitudes and susceptibility to persuasion, and thus the outcomes of an experiment. For example, a group of persons who were administered a pretest dealing with hostility and then viewed a film dealing with prejudice, may have a different reaction than those who did not take the pretest first. In this case, the testing itself would have an effect unrelated to the exposure of the movie.[3]

In the everyday world, tests are administered under different circumstances. Testing is considered a usual event in settings such as educational institutions. However, in a business setting, testing may be viewed suspiciously, increasing the likelihood that factors such as surprise or stress may play a role in the manner that questions on a pretest are answered. The uniqueness of these environments affect the ability of the researcher to draw parallels to wider audiences, and thus limit conclusions that may be drawn from experimental results. Campbell and Stanley explain this phenomenon by noting that "the effect of the pretest upon X as it restricts external validity is of course a function of the extent to which such repeated measurements are characteristic of the universe to which one wants to generalize."[3(p18)] In environments where testing is an unusual occurrence, the Testing-X interaction comes into play, and generalizability may be enhanced by using comparison groups who are *not* administered a pretest.

How does the Testing-X interaction apply to the Internet? As an example, assume a researcher wishes to know how persons with different types of chronic disease cope with their illness. Suppose participants were directed to a Web site about "Coping with Your Illness," and were administered a pretest asking questions about coping, such as "I find that I can't handle everyday chores," and "I feel overwhelmed a lot." Once the survey was completed, they were asked to join an Internet support group led by a person who was an expert in coping with chronic illness. Several weeks later, they were then administered a posttest to measure the effectiveness of the intervention. How would the researcher know if the scores were affected by the questions in the pretest? Did the questions perhaps affect their stress level, or their belief in their ability to cope? The researcher will note that the Testing-X interaction, albeit with differences in implementation of testing methods, has similar characteristics when examined in relation to Internet research and traditional venues.

Suppose the researcher redesigned the experiment to include 4 randomized groups, one who took the posttest only, one who were exposed to the support group, then took the posttest, one who took the pretest and posttest only, and one who took the pretest, were exposed to the support group, then took the posttest. By comparing the groups on their posttest scores, a comparison would be able to be made on the presence of an interaction between those who may have been exposed to the "giveaway" content of the pretest.[3] What is the impact on generalizability using this design? By using the unpretested group, the researcher is able to see if there is a difference in the characteristics of those who were exposed to the posttest, versus the audience to which one wants to generalize, namely persons on the Internet with chronic illness. This concept is more thoroughly explored under the Solomon Four-Group Design.[3]

The interaction of selection and X

Selection is a second threat to internal validity that may affect generalizability. Selection refers to the "biases resulting in differential selection of respondents for the comparison groups."[3(p5)] In research design,

selection of subjects who are representative of the population is crucial, so that if a difference exists between control and experimental groups, the effect would be because of the exposure. The question of any difference being applicable only to the unique population from which the groups were jointly selected remains a concern. As difficulty in procuring subjects increases, it becomes more likely that nonrepresentativeness of the population will increase because of a selection-specificity effect. In work settings, this could be in the form of effects related to motivational participation factors, such as staff morale, fear of inspection/disclosure, or zeal for improvement. For example, if 7 facilities were approached to conduct a study, and only one agreed, the difference in characteristics between the agreer and the nonagreers should be examined. If the agreeing agency had a characteristic not present in the nonagreers, for instance a public service orientation, generalizability could be threatened. The same holds true for groups of individuals studied; is there a difference between the groups, for instance mean age, or outside participation in activities? Do those who participate feel a need to be important? Do they have preexisting "hot buttons" for certain causes that would lead them to participate? Do they have more time because of social or economic circumstances? Also, the researcher must remember that participants are often selected on the basis of cooperativeness and accessibility. Thus, a sampling bias can occur if the phenomenon is not demonstrated over a wide variety of conditions, and generalizability may be enhanced by replication with diverse heterogeneous groups.[3]

On the Internet, it is possible to assess the *interaction of selection and X* by randomizing participants into 2 groups, then testing one before an intervention, and one afterwards. For instance, suppose a market researcher wished to study the effects of students reading about new bibliographic software on their propensity to later purchase it. One group of participants could be assessed on their attitudes about bibliographic software, and then be directed to a Web site that contained positive reviews about the software. The second group would read the review first, and then asked their attitudes. This Separate-Sample Pretest–Posttest design, as well as variations of it, would allow the researcher to control for the *main effect* of testing, as well as the *interaction* of testing and X. Campbell and Stanley note that these designs "are apt to be superior in external validity to true experiments . . . [they] put so little demand upon the participants for cooperation, for being in certain places at certain times, etc., that representative sampling from populations specified in advance can be employed."[3(p54)] As can be seen, these types of designs can be readily applied to the Internet, which by definition comprises persons in *different* places at *different* times.

Other interactions with X

Although not all threats to internal validity will interact with X to cause a threat to generalizability, one, several, or all of these threats in combination may interact synergistically, and must be considered. These include history, maturation, etc. They will each be reviewed independently, and then in combination. Two threats to internal validity, regression and mortality, do not interact with X to cause threats to generalizability, and thus will not be covered here. *Statistical regression*, recognized as "operating where groups have been selected on the basis of their extreme scores,"[3(p5)] does not interact with X. Similarly, *experimental mortality*, or the "differential loss of respondents from the comparison groups,"[3(p5)] does not interact with X, and is instead a product of X.[3]

History

According to Campbell and Stanley, *history* is defined as "the specific events occurring between the first and the second measurement in addition to the experimental variable."[3(p5)] The *interaction of history and X* may be

time-dependent, ie, current events at the time of the testing may have an influence on the experimental results. Generalizability is enhanced by replication across time periods and settings. For example, types of persons who would volunteer to respond to a study regarding their adherence to a low-sodium diet might be different than at other times of the year on the Internet. For instance, is there a possibility that those who do not follow their diet are more sedentary, and thus might be more available to volunteer in the summer, when those who follow their diet might be more active? If the researcher attempted to generalize findings from the summer to all persons on the Internet who are on low-sodium diets, inferences may not be based on the true population. By varying time periods of the study, the researcher is increasingly likely to get a more accurate estimate of the population, and thus generalizability. In addition, the pace of "time" on the Internet has been characterized as occurring faster than "real" time. Because of the stimulation of technology, events are perceived as "happening" faster; persons may feel as if much more time has elapsed than in truth has actually occurred. For example, a "net-year" may only be 3 "real" months.[11] This phenomenon may have implications for participants in studies dealing indirectly or directly with the effects of the Internet itself. For example, exposure to computer viruses, hoaxes, etc may influence the generalizability of a survey on attitudes toward computers in populations that are Internet users, versus those who are not.

Maturation

Maturation is defined as "processes within the respondents operating as a function of the passage of time per se (not specific to particular events), including growing older, growing hungrier, growing more tired, and the like."[3(p5)] The *interaction of maturation and X* can be a factor in external validity. Psychological or biological processes that systematically change over time, such as boredom or gaining experience, may occur in the research participants. If these processes are not related to specific external events, there may be an influence in treatment outcomes, and thus impact generalizability to a population beyond that which is currently under study.[3] For instance, suppose a researcher wishes to administer a common psychological test on the Internet. If participants were experienced in this type of testing, for instance those receiving long-term psychotherapy, the generalizability of findings to other populations may be affected.

Instrumentation

Campbell and Stanley define *instrumentation* as a phenomenon "in which changes in the calibration of a measuring instrument or changes in the observers or scorers used may produce changes in the obtained measurements"[3(p5)] Specific effects can occur in relation to the instrumentation of the study to *interact with X*. Although instrumentation is considered a threat to *internal* validity, specific tests, observers, or measuring instruments may interact to affect experimental results. Generalizability is enhanced when multiple observers or interviewers are used across treatments, and the effects can be measured directly.[3] For example, suppose an educational researcher is interested in studying the effects of an online Windows Tips computer class on confidence in computer ability. If one instructor taught the class, versus if there were 2 classes, each with different instructors, generalizability would be less because it would be unknown whether the instructor or the class content was responsible for any changes in confidence. Thus, by sampling participants from both classes in this case, the researcher would be more confident in suggesting that study findings might be applied to students in other Windows Tips classes, and depending upon the strength of other factors in the study, such as the demographic characteristics of the participants, possibly to other types of computer classes as well.

Interaction of selection and maturation

Threats to generalizability may also occur in combination. For instance, history and maturation may interact. Elderly persons who are normally homebound in the winter may be less likely to use the Internet in the summer, when it is warmer, and there are less obstacles to going outdoors. By assessing potential participants for their age, as was noted in Table 1, and activity status by season, the researcher would be able to determine if the group is representative of typical elderly people, and may choose to sample at different times of the year in order to increase potential generalizability to a wider range of elderly persons.

Multiple treatment interference

Threats to generalizability can occur through multiple treatment interference. Multiple treatment interference takes 2 forms. In the first form, participants are administered more than one treatment. If participants receive both treatments simultaneously, there may be an interaction between them. If the treatments are administered sequentially, there may be a crossover effect, in which the earlier treatment affects the later one. In the second form of multiple treatment interference, subjects participate in more than one study, or multiple studies. This experience in the research process, particularly if debriefings or explanations of previous studies occurred, may affect participants' behavior in the current study.[10]

On the Internet, multiple treatment interferences may be similar to those of a physically present population. For instance, suppose a researcher is planning a study, and in reviewing the literature, learns of Web sites that were cooperative in posting announcements of studies. If the newest study is posted on one of those Web sites, depending on the number of returning users, there would be a possibility that the researcher may be recruiting "experienced" participants. In turn, this would impact the subsequent generalizability of the study, for the researcher would be unable to determine if the subjects' previous research experience was a factor.

Reactive arrangements

The effects of participants' knowledge that they are being studied, or the Hawthorne Effect,[12,13] may have an impact on their reaction to the study. An individual may react to an experiment based on its stimulus value, as well as implicit hints of the purpose of the study. The latter cognitive *process* may be contrasted to the Testing-X threat to external validity, where the *content* of the test is the issue of interest. The effect of this reaction will be embedded into the posttest results. Campbell and Stanley note that if the subsequent attitudes engendered by the novelty of the situation are unrepresentative of the setting being studied, they may be "qualifiers of the effect of X, seriously hampering generalization."[3(p20)] Increased generalizability is attempted by setting experimental conditions to be as usual and ordinary as possible, as "the more obvious the connection between the experimental treatment and the posttest content, the more likely this effect becomes."[3(p21)] In addition, in group settings, focusing interventions on individual participants without knowledge of the others, and careful randomization of exposure groups may be effective strategies. For instance, one class could be given one version of a test, and another class given a second version. Familiarity of the researcher to the subjects must also be addressed. Optimally, a balance must be achieved between utilizing researchers who are "outsiders" to the setting (ie, are unknown to the study participants), and achieving rigorous but not wholly applicable results, versus using "insiders" (or those who typically belong in the setting), and potentially resulting in less rigor, but yielding more applicable results. Moreover, in educational settings, other strategies to decrease threats to external validity because of reactive arrangements include timing test administration during plausible times and embedding questions during routine examinations.

On the Internet, it is difficult to decrease the possibility of reactive arrangements. Because the venue is primarily not educational, test-taking is not a usual event, and could be considered not commonplace, or "artificial." The exception to this novelty might be online classes, the Internet version of an educational setting, where test-taking might be expected. For instance, a posttest in an online class about sociology might contain embedded questions that measure attitudes about families. There would be constraints in generalizability in that results would be difficult to apply to students in a traditional sociology class. However, ethical implications arise in testing individuals without their knowledge. Human rights and informed consent considerations must be strictly adhered to, and participants must be given information on the reason for the study, the risks and benefits of participating, and the freedom to withdraw from participation at any time.[14] Therefore, in this instance subjects would need to be informed that they will be answering questions about attitudes toward families embedded in their posttest. Those electing not to answer the questions would need to be given a posttest that does not contain the questions about attitudes toward families.

Furthermore, participants' knowledge that they are being studied is not only crucial from an ethical standpoint, but also from a legal perspective. In areas of the Internet where health care issues are addressed, governmental regulations regarding the privacy of personal and medical information must be taken into account. Section 264 of the Health Insurance Portability and Accountability Act (HIPAA), enacted into law in 1996,[15] calls for the establishment of regulations related to medical privacy. As of this writing, H.R. 1941, the Health Information Privacy Act (HIPA),[16] which explicitly outlines the rights of persons to have the minimum amount of information necessary for treatment and payment to be transferred, electronically or otherwise, between those who are involved in the individual's care, is still in the legislature. The researcher must remember that because of this legislation and its aftermath, a new awareness of privacy issues has developed among Internet users.[17] Internet health sites that lack adequate privacy policies may be noticeable to even a novice Internet user.[18] This may raise questions in those who might participate in online health research, and result in different groups of persons being willing to participate in particular sites. This may be particularly true of those with stigmatic diagnoses.[19] Thus, the momentum toward privacy created by HIPA may impact potential recruitment of participants, and ultimately, generalizability.

A proposed sixth factor: Extraneous threats related to networked effects

The increasingly ubiquitous nature of the Internet may contribute to issues of generalizability. For each research study, not only must the above 5 factors be examined, but also extraneous threats due to *networked* effects. The *linked* nature of the Internet provides for possibilities that may not be in the immediate consciousness of the researcher, and must be explicitly examined. The synergistic nature of networks and links may cause threats to generalizability to occur alone or in combination. Networked threats to generalizability may be grouped into 4 categories: (a) co-occurring interferences to testing, (b) self-selection mechanisms, (c) electronic group environments, and (d) cultural differences. They are outlined below.

Co-occurring interferences to testing

Co-occurring interferences to testing may also occur simultaneously on the Internet, *affecting exposure to stimuli independent of the experiment*, both *before* and *during* pretests and posttests, providing a unique form of the Testing-X interaction. Unintended networked entry threats may impact research participants not only preexposure but also *during* exposure. There is no certainty that a participant filling out a survey, or reading material of the exposure etc, is not *multitasking*. For example, a subject may be simultaneously viewing other sites, or reading incoming e-mail. In addition, the participant may be viewing something in another window that

affects the testing. For instance, a respondent may be reading about depression while taking a pretest on depression. Moreover, the ability to multitask on the Internet offers numerous exit paths from a research site. Participants who are multitasking may see something more interesting in another window of their browser, and not complete the study, especially if it is time-consuming. In addition, because the researcher is not physically present, there is no potential embarrassment in not being a "good participant." The result is a group of persons who are focused, and interested enough in the research, or in contributing, to persevere. Thus, a sample may represent a narrow slice of the population it is intended to represent, and depending on the nature of the study, seriously impact generalizability. There is little the researcher can do about this other than to emphasize to participants the importance of their contribution to the study, or offer an incentive such as a monetary acknowledgment for completion. However, if participants are compensated, the researcher must be aware that the nature of participation will be changed, and may attract those with more socioeconomic need.

In addition, co-occurring interferences with testing can occur because of environmental factors. The nature of the testing environment may influence participants' responses. For instance, responses may be different if the test is taken in a lab or at home. Moreover, technological factors such as browser software, and other hardware and software configurations, as well as the speed of different computers may result in presentation of the same test in a completely different manner, ie, the equivalence of the instruments may be affected. If frustration due to presentation or speed is a result, participants who eventually complete a test represent those who are more tolerant, and not the entire population.[20]

Self-selection mechanisms

By examining entry mechanisms to studies, the researcher will note that a magnified process of self-selection occurs that is beyond the traditional, non-Internet research study. The characteristics of persons who ultimately become aware of a research destination on the Internet may create a very narrow group of participants who are particularly interested in a specific topic, which in turn may affect heterogeneity, and ultimately, generalizability. By envisioning the process by which a particular user reaches a research site, the researcher can assess the characteristics of persons who may arrive there, and determine the potential for the heterogeneity, or lack thereof, of the users. Very few Web sites stand alone, ie, they do not have links to other Web sites outside of their location, and are thus subject to lack of control of the research milieu.

The pervasive nature of search engines and "clickable links" provide multiple access points to an Internet location. Entry to a site may occur through a self-selected search engine "hit," from a site originally found in a search engine, or from a link in an e-mail or a newsgroup; the commonality is persons who have specifically selected that particular topic, in contrast to all that is available on the Internet at that particular time. The reductive process of "drilling down" and eliminating what is *not* wanted results in a pool of highly selected, interested users to a particular location. For example, a researcher studying the entrance pathways to a political site for a candidate may find that many people are entering through links from (i) the National Democratic Party (ii) the Yahoo search engine, or (iii) e-mail. Is there a difference in these type of users, versus a more random group of Internet users who may have not particularly intended to seek out political sites that day? The phenomenon was demonstrated in health care by Taubenheim,[21] in a study on the use of the Internet to search for information arthritis and musculoskeletal diseases. Of the 707 respondents, 51% learned of the study through a participating organization's Web site, while 22% linked to it through another site. In addition, 41 different topics related to arthritis and musculoskeletal skin diseases were searched for by respondents, demonstrating how the pool of participants was gradually narrowed down to a selective audience.

Self-selection mechanisms on the Internet have important implications for generalizability because although the researcher may examine demographic characteristics of the sample, and determine that they are similar in relation to a larger population, there may be characteristics related to self-selection that are indeterminable, and an apparent likelihood between groups may not be so. In addition, potential threats to generalizability from multiple completions of tests and mischievous responding must be considered. Although submission of duplicate submissions by participants can be tempered somewhat by password protection as well as an examination of server logs, they could affect the independence of observations, and thus, generalizability.[20]

Electronic group environments

Threats to generalizability can occur anywhere on the Internet where persons gather. These threats to generalizability may be more complex than those of reactive arrangements alone because the multiple potential for communication exchanges between study participants may influence prior knowledge of the study. In both synchronous, where communication occurs in real time, and asynchronous communication, where interactions occur at different times, the presence of others may impact the research environment. For instance, groups of people with a common interest may know each other from a newsgroup, a discussion list, and a chat area. In each area, they may be affected by the presence of others, or hear of what the study was like. They may absorb information without interacting, or others being aware of their presence; in Internet parlance this is known as "lurking." Especially in electronic gathering venues that deal with the same topic as the research, potential participants may mention their involvement in the study, and dialogue may occur about the nature or expectations of the study. For example, participants in an asthma chat room may tell one another about an asthma self-efficacy study and discuss how they feel about their asthma. If any of these listeners then go on to participate in the actual study, pretest results may be influenced by this discussion, and thus generalizability of findings. Table 2 delineates extraneous threats to generalizability related to networked effects, and outlines various forms of asynchronous and synchronous electronic group environments that the researcher must assess.

Table 2. Extraneous threats to generalizability related to networked effects

Threats	Example
Co-occurring interferences to testing	Multitasking Environmental factors Technological factors
Self-selection mechanisms	"Drilling down" process to reach a site
Electronic group environments	Asynchronous venues • Chat rooms • Instant messenger conversations Synchronous venues • E-mail • Electronic discussion lists • Newsgroups • Web-based discussion boards
Cultural differences	Language Mores Health systems Educational levels

Cultural differences

In addition to the above threats to generalizability, cultural differences may also play a role because of the networked effects on the Internet. The researcher must be cognizant of potential assumptions related to culture, as well as potential interpretations of the research by persons in other cultures. Multilingual persons may interact on the Internet in more than one language, and as noted by the many questions posed earlier in this

article, the effects of this on the research are unknown, but must be considered. Cultural mores may also be a factor, as well as meanings of typical health care in countries with nationalized systems, and fewer resources. In addition, if the researcher is assessing level of education, not all countries have educational systems similar to the United States. If the researcher wishes to generalize to high school graduates, for example, European countries sometimes have "forms" as opposed to grade levels, which are not necessarily equivalent to US grade levels. Although persons from the United States are by far the dominant "presence" on the Internet,[22] potential cultural aspects on generalizability must be assessed.

Thus, threats to generalizability due to extraneous networked effects may be noted by examining co-occurring interferences to testing, self-selection mechanisms, electronic group environments, and cultural differences. In summary, this section of the article has outlined threats to generalizability that may occur as a result of the interaction of selection and X, the interaction of testing and X, other interactions with X, multiple treatment interference, reactive arrangements, and threats due to extraneous networked effects. Next, generalizability in qualitative research will be examined. The methodologies of qualitative research present a different set of issues not present with quantitative studies.

ISSUES IN QUALITATIVE RESEARCH

Generalizability of qualitative research presents unique issues when applied to the Internet, particularly when contrasted to quantitative research. The major differences between the 2 will be explained, and then several issues of generalizability in relation to qualitative research will be examined. Finally, an overview of emergent topics in qualitative generalizability related to Internet research will be considered.

In the quantitative tradition, the goal of research is deduction, or theory-testing, whereas qualitative research is induction, or theory-building. The two traditions can be viewed on a continuum.[23] In the deductive process, specific hypotheses are tested on the basis of a general theory. In the inductive process, specific data are reduced into their common components in order to build a general theory. In quantitative research, the data are primarily objective, and phenomena are validated by empirical methods. In contrast, qualitative research is distinguished by its holistic study of the human experience, or the subjective study of persons' realities.[24,25] In quantitative research, the validity and reliability of instruments used to measure phenomena are typically established before the commencement of a study. This is distinguished from qualitative research, where the researcher *is* the instrument through which data flow.[26]

Because qualitative research focuses more on "understanding particulars," as opposed to "generalizing to universals" as is true with quantitative research, elimination of threats to generalizability is often less possible.[27(p296)] However, among researchers, there is lack of consensus as to whether the concept of generalizability, or even validity itself,[28] is fitting for qualitative research, and if so, how it should even be considered. These differences of opinion can be categorized as either interparadigmatic, or intraparadigmatic.

Interparadigmatic differences regarding generalizability in qualitative research have occurred because of philosophical traditions. Quantitatively oriented researchers have noted that in qualitative research, because there is a typically small number of participants, there is an absence of randomness, and research results are applicable only to the individuals studied.[29] Morse elaborates on this phenomenon:

> The generalizability of research findings has long been considered the prerogative of quantitative research. The selection of an adequate and random sample and the comparison of the study population with the sample have been techniques that permit the transference of the quantitative study results to the study population and to similar populations. Using this standard, qualitative research, with small and purposefully selected samples, has

been considered nonrepresentative of the population and its findings not generalizable.[30(p5)]

However, others feel that qualitative research should stand on its own merit, and should not be considered in empirical terms. Beck[31] proposes that the concept of *fittingness* in qualitative research is analogous to that of generalizability in quantitative research. It is described by Guba and Lincoln as "how well the working hypotheses or propositions fit into a context other than the one from which they were generated."[32(p264)] Morse does not use this terminology, although she describes in nonquantitative terminology how generalizability may be enhanced:

> In qualitative research, each participant in the relatively small sample has been selected purposefully for the contribution he or she can make toward the emerging theory. It is this selecting that ensures that the theory is comprehensive, complete, saturated, and accounts for negative cases . . . The theory also is applicable beyond this immediate group and is applicable to all similar situations, questions, and problems, regardless of the comparability of the demographic composition of the groups.[30(p5)]

By using the example of privacy violations of elderly males in a nursing home, Morse demonstrates how findings would be applicable to female psychiatric patients, stating that "The knowledge gained is not limited to demographic variables; it is the fit of the topic or the comparability of the problem that is of concern. Recall it is the knowledge that is generalized."[30(p6)] Thus, in regard to generalizability, there is a difference between qualitative and quantitative research. In quantitative research, small nonrandom sample sizes are worrisome because of their nonrepresentativeness.[3] However, in qualitative research, fittingness, the qualitative equivalent of generalizability, may be assessed on 5 dimensions: (1) the typicality of the informants and their responses; (2) the representativeness of the data as a whole; (3) theoretical sampling, which resulted in a range of informants experiencing the phenomena under study; (4) data, which did not appear to be more similar or congruent than they really were; and (5) study results that fit the data from which they were generated.[33]

As contrasted to these interparadigmatic issues in qualitative generalizability/fittingness, intraparadigmatic issues have been noted by Maxwell. He distinguishes between internal and external generalizability, distinguishing them as "*within* the community, group, or institution studied to persons, events, and settings that were not directly observed or interviewed, and generalizability *to* other, communities groups, or institutions [emphasis added]."[27(p293)] Generalizability has not been explicitly addressed in relation to different qualitative methods. However, Beck,[33] in highlighting general validity issues between 3 phenomenological research methods, demonstrates a lack of consensus on the meaning of validity *within* the phenomenological movement. For instance, using Colaizzi's phenomenological method,[34] the researcher achieves final validation by returning to the participant, whereas VanKaam[35] requires expert judges to reach intersubjective agreement, and Georgi[36] calls for the researcher to self-determine validity. Although the methods do not expressly address generalizability, she explains that when using Lincoln and Guba's criteria for qualitative rigor,[37] the researcher must determine the applicability of findings to other contexts or subjects.

How does qualitative generalizability apply to the Internet? At best, this depends on the context of the study. For example, suppose a researcher intended to study the lived experience of persons with heart disease who use online support groups. As contrasted with quantitative research, where the generalizable population would be limited, if one used Morse's criteria,[30] the group would be much larger, as the context could be applied to a number of groups on the basis of the comparability of the problem. For instance, the findings could be applied to persons who suffer from a chronic disease such as lupus, or those who are isolated, and use online support groups as a social mechanism. In addition,

population differences could be transcended because of more universality of experience. For instance, persons in the United States and the United Kingdom would represent different groups in a quantitative study, yet the *experience* of heart disease may be comparable.

In addition, there are other emergent issues related to the nature of the Internet medium, and its comparison to traditional study methods. As with quantitative research, the storage of observations and field notes must be addressed from privacy as well as ethical and legal perspective, as explained earlier in this article. Particularly in qualitative research, privacy of message content from Internet support groups must be.[38] If not, it may influence participation in studies, and potential generalizability. In addition, differences between the written word and the spoken word must be addressed. Does posing the same question on the Internet, and using face-to-face interviews yield the same results and hermeneutic richness?[39] If so, there could be an impact on generalizability. Normally unrevealed phenomena, such as the ambiance of the environment, or facial expressions, may have an impact on participants' responses, and thus generalizability to larger populations.

Qualitative research can also assist in highlighting differences and similarities between Internet and traditional formats, and help explain potential issues in generalizability. For instance, Scherrer-Bannerman and colleagues[40] compared the impact of Web-based education and support for 72 patients awaiting cardiac surgery, and found that Web-based support was superior in increasing social support and decreasing anxiety, but that hindering factors were lack of confidence in computer ability, and lack of comfort in communicating electronically with physicians and nurses. Similarities cited between the formats were reassurance of having information, and assurance of survival after surgery and the ability to continue with life.

In summary, interparadigmatic as well as intraparadigmatic differences are present when looking at generalizability in qualitative research. In spite of differences from the empirical paradigm, representativeness due to fit of the topic and comparability of the problem is applicable to qualitative generalizability; knowledge gained overshadows the essential issues related to sample size that must be considered in quantitative research. These differences can be applied to the Internet, particularly to universality of experience in far-reaching populations.

SUMMARY AND IMPLICATIONS

The implications of generalizability are wide and far-reaching. On the Internet, generalizability must be considered from both a quantitative and qualitative standpoint. Although Campbell and Stanley 3 provide a guide for addressing quantitative research, continual assessment of emerging issues must occur as new facets of Internet technology become amenable to research. Tools such as wireless devices will present new challenges in access to participants, and thus, generalizability. As changes in online populations occur, the researcher must determine the impact of such shifts on the generalizability of samples and intended samples based on research objectives, and modify methodology accordingly. Traditional as well as networked threats to generalizability must be taken into account in both quantitative and qualitative research paradigms.

Moreover, on a policy level, changes are beginning to occur. The Science Panel on Interactive Communication and Health of US Department of Health and Human Services[41] has begun to take notice of the validity of Internet research, emphasizing the need for ongoing assessment of factors that may impact generalizability in relation to the Internet. In order to keep current of issues in generalizability, the researcher must make no assumptions about the characteristics of online populations, and must be continually alert to evolving changes in technology and user interfaces that may impact the future of both quantitative and qualitative Internet research.

REFERENCES

1. Jones S, ed. *Doing Internet Research: Critical Issues and Methods for Examining the Net*. Thousand Oaks, Calif: Sage; 1999.
2. Birnbaum MH, ed. *Psychological Experiments on the Internet*. San Diego, Calif: Academic Press; 2000.
3. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally; 1963.
4. Reese WL. *Dictionary of Philosophy and Religion: Eastern and Western Thought*. Atlantic Highlands, NJ: Humanities Press; 1996.
5. Cook TD, Campbell DT. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Chicago: Rand McNally; 1979.
6. Trochim W. *The Research Methods Knowledge Base*. 2nd ed. Ithaca, NY: Cornell Custom Publishing; 1999.
7. *Webster's New World Dictionary and Thesaurus*. M. Agnes, ed. Principal thesaurus text by Laird C. New York: Simon & Schuster; 1996.
8. The Pew Internet and American Life Project. *Tracking Online Life: How Women Use the Internet to Cultivate Relationships With Family and Friends*. Washington, DC: The Pew Internet and American Life Project; 2000. Available at: http://www.pewinternet.org/reports/toc.asp?Report=11. Accessed April 14, 2003.
9. US Census Bureau. *American Fact Finder Detailed Tables: PCT12. SEX BY AGE [209]—Universe: Total Population. Data Set: Census 2000 Summary File 1 (SF 1) 100-Percent Data United States*. Washington, DC: US Census Bureau; 2000. Available at: http://factfinder.census.gov/servlet/DTTable . . . geo_id = 01000US&mt_name = DEC_2000_SF1_U_PCT012. Accessed April 14, 2003.
10. Pedhazur EJ, Schmelkin LP. *Measurement, Design and Analysis: An Integrated Approach*. Hillsdale, NJ: Lawrence Erlbaum; 1991.
11. Public Broadcasting System (PBS). *Nerds 2.01: A Brief History of the Internet* [videotape]. Alexandria, Va: Public Broadcasting System; 1999.
12. Bracht GH, Glass GV. The external validity of experiments. *Am Educ Res J*. 1968;5:437–474.
13. Buckwalter KC, Maas ML. True experimental designs. In: Brink PJ, Wood MJ, eds. *Advanced Design in Nursing Research*. Newbury Park, Calif: Sage; 1989:27–56.
14. US Department of Health and Human Services. *Code of Federal Regulations, Title 45, Part 46: Public Welfare-Protection of Human Subjects*. Washington, DC: US Dept of Health and Human Services, National Institutes of Health Office for Protection From Research Risks; 1991. Available at: http://ohrp.osophs.dhhs.gov/humansubjects/guidance/45cfr46.htm. Accessed April 14, 2003.
15. US Congress. *Health Insurance Portability and Accountability Act of 1996*. Public Law No. 104–191. Washington, DC: US Government; 1996. Available at: http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=104_cong_public_laws&docid=f:publ191.104.pdf. Accessed April 14, 2003.
16. US House of Representatives. *Health Information Privacy Act*. Introduced in the House May 25, 1999. Washington, DC: US Government; 1999. Available at: http://frwebgate.access.gpo.gov/cgi-bin/getpage.cgi?dbname=1999_record&page=E1086&position=all. Accessed April 14, 2003
17. LaCoursiere SP. Development and evaluation of an instrument to measure Attitudes Toward Online Health Care (ATOHC). In: Proceedings of the Eastern Nursing Research Society 13th Annual Scientific Sessions. Atlantic City, NJ; April 1–3, 2001.
18. Harris Interactive and the Internet Healthcare Coalition. *Ethical Issues Awareness of e-Health Site Users vs. Developers: A Different Kind of Digital Divide*? Available at: http://www.ihealthcoalition.org/content/Harris_survey2000.pdf. Accessed April 14, 2003.
19. Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. *Am Psychol*. 2000;55:205–217.
20. Buchanan T, Smith JL. Using the Internet for psychological research: personality testing on the world wide web. *Br J Psychol*. 1999;90:125–144.
21. Taubenheim AM. *Conducting Survey Research on the Internet: Use of the Internet to Search for Information on Arthritis and Musculoskeletal and Skin Diseases* [dissertation]. College Park, Md: University of Maryland; 1999.
22. Commerce Net. *Worldwide Internet Population*. Available at: http://www.commerce.net/research/stats/wwstats.html. Accessed April 14, 2003.
23. Munhall PL. Epistemology in nursing. In: Munhall PL, Boyd CO, eds. *Nursing Research: A Qualitative Perspective*. 2nd ed. New York: National League for Nursing Press; 1993:39–65. Pub. No. 19–2535.
24. Eisner EW. *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*. New York: Macmillan; 1991.
25. Boyd CO. Philosophical foundations of qualitative research. In: Munhall PL, Boyd CO, eds. *Nursing Research: A Qualitative Perspective*. 2nd ed. New York: National League for Nursing Press; 1993:66–93. Pub. No. 19–2535.
26. Boyd CO, Munhall PL. Qualitative research proposals and reports. In: Munhall PL, Boyd CO, eds. *Nursing Research: A Qualitative Perspective*. 2nd ed. New York: National League for Nursing Press; 1993: 424–453. Pub. No. 19–2535.
27. Maxwell JA. Understanding and validity in

qualitative research. *Harvard Ed Rev*. 1992;62(3): 279–299.
28. Salner M. Validity in human science research. *Saybrook Rev*. 1986;6:107–130.
29. Hycner RH. Some guidelines for the phenomenological analysis of interview data. *Hum Stud*. 1985;8:279–303.
30. Morse JM. Qualitative generalizability. *Qual Health Res*. 1999;9:5–6.
31. Beck CT. Qualitative research: the evaluation of its credibility, fittingness, and auditability. *West J Nurs Res*. 1993;15:263–266.
32. Guba E, Lincoln Y. *Effective Evaluation*. San Francisco: Jossey-Bass; 1981. Cited in Beck CT, ed. Qualitative research: the evaluation of its credibility, fittingness, and auditability. *West J Nurs Res*. 1993;15:263–266.
33. Beck CT. Reliability and validity issues in phenomenological research. *West J Nurs Res*. 1994;16:254–267.
34. Colaizzi P. Psychological research as the phenomenologist sees it. In: Valle R, King M, eds. *Existential-Phenomenological Alternatives for Psychology*. New York: Oxford University Press; 1978:48–71.
35. VanKaam A. *Existential Foundations of Psychology*. Pittsburgh: Duquesne University Press; 1966.
36. Georgi A. *Phenomenology and Psychological Research*. Pittsburgh: Duquesne University Press; 1985.
37. Lincoln Y, Guba E. *Naturalistic Inquiry*. Newbury Park, Calif: Sage; 1985. Cited in Beck CT, ed. Reliability and validity issues in phenomenological research. *West J Nurs Res*. 1994;16(3):254–267.
38. Klemm P, Nolan MT. Internet cancer support groups: legal and ethical issues for nurse researchers. *Oncol Nurs Forum*. 1998;25:673–676.
39. Ngwenyama OK, Lee AS. Communication richness in electronic mail: critical social theory and the contextuality of meaning. *MIS Q*. 1997;21(2):145–167.
40. Scherrrer-Bannerman A, Fofonoff D, Minshall D, et al. Web-based education and support for patients on the cardiac surgery waiting list. *J Telemed Telecare*. 2000;6(suppl 2):72–74.
41. US Dept of Health and Human Services, Office of Disease Prevention and Health Promotion, Science Panel on Interactive Communication and Health. Eng TR, Gustafson DH, eds. *Wired for Health and Well-Being: The Emergence of Interactive Health Communication*. Washington, DC: US Dept of Health and Human Services, U.S. Government Printing Office; 1999.