

Sampling Designs in Qualitative Research: Making the Sampling Process More Public

Anthony J. Onwuegbuzie

Sam Houston State University, Huntsville, Texas

Nancy L. Leech

University of Colorado at Denver, Denver, Colorado

The purpose of this paper is to provide a typology of sampling designs for qualitative researchers. We introduce the following sampling strategies: (a) parallel sampling designs, which represent a body of sampling strategies that facilitate credible comparisons of two or more different subgroups that are extracted from the same levels of study; (b) nested sampling designs, which are sampling strategies that facilitate credible comparisons of two or more members of the same subgroup, wherein one or more members of the subgroup represent a sub-sample of the full sample; and (c) multilevel sampling designs, which represent sampling strategies that facilitate credible comparisons of two or more subgroups that are extracted from different levels of study. Key Words: Qualitative Research, Sampling Designs, Random Sampling, Purposive Sampling, and Sample Size

Setting the Scene

According to Denzin and Lincoln (2005), qualitative researchers must confront three crises; representation, legitimation, and praxis. The crisis of representation refers to the difficulty for qualitative researchers in adequately capturing lived experiences. As noted by Denzin and Lincoln,

Such experience, it is argued, is created in the social text written by the researcher. This is the representational crisis. It confronts the inescapable problem of representation, but does so within a framework that makes the direct link between experience and text problematic. (p. 19)

Further, according to Denzin and Lincoln (2005), the crisis of representation asks whether qualitative researchers can use text to represent authentically the experience of the "Other" (p. 21). The crisis of legitimation refers to "a serious rethinking of such terms as *validity*, *generalizability*, and *reliability*, terms already retheorized in postpositivist..., constructivist-naturalistic..., feminist..., interpretive..., poststructural..., and critical...discourses" (Denzin & Lincoln, p. 19) [*italics in original*]. Finally, the crisis of praxis leads qualitative researchers to ask, "how are qualitative studies to be evaluated in the contemporary, poststructural moment?" (Denzin & Lincoln, pp. 19-20).

The crises of representation, legitimation, and praxis threaten qualitative researchers' ability to extract meaning from their data. As noted by Onwuegbuzie and Leech (2004a),

In particular, lack of representation means that the evaluator has not adequately captured the data. Lack of legitimation means that the extent to which the data have been captured has not been adequately assessed, or that any such assessment has not provided support for legitimation. Thus, the significance of findings in qualitative research is affected by these crises. (p. 778)

In an attempt to address these crises and to prevent "the naturalistic approach... [from being] tarred with the brush of 'sloppy research'" (Guba, 1981, p. 90), in recent years, there has been increased focus on rigor in qualitative research, where rigor is defined as the goal of making "data and explanatory schemes as public and replicable as possible" (Denzin, 1978, p. 7). More specifically, recent attempts have been made to make the research process more public (cf. Anfara, Brown, & Mangione, 2002). In particular, qualitative methodologists have provided frameworks for making qualitative data analyses more explicit (Anfara et al.; Constan, 1992), so that qualitative studies promote "openness on the grounds of refutability and freedom from bias" (Anfara et al., p. 28).

In contrast, scant discussion has taken place vis-à-vis sampling in qualitative research. Indeed, using the keywords "qualitative research" and "sampling," as well as "qualitative research" and "sample size," a review of the most prominent academic literature databases (e.g., ERIC, PsycINFO) yielded only seven published journal articles (i.e., Crowley, 1994; Curtis, Gesler, Smith, & Washburn, 2000; Jones, 2002; Merriam, 1995; Onwuegbuzie & Leech, 2004b, 2005b; Sandelowski, 1995) that discussed the issue of sampling and/or sample size in qualitative research. Additionally, Onwuegbuzie and Leech (2005a), Collins, Onwuegbuzie, and Jiao (2006, 2007), and Teddlie and Yu (2007) have added to the body of literature in this area. All of these articles have focused on the issue of sample size and/or sampling schemes. Although these concepts are extremely important in interpretivist research, none of these articles provide a superordinate concept of sampling designs. For the purposes of the present essay, we distinguish between sampling schemes and sampling designs. We define sampling schemes as specific techniques that are utilized to select units (e.g., people, groups, subgroups, situations, events). In contrast, as do Onwuegbuzie and Collins (2007), we define sampling designs as representing the framework within which the sampling occurs, comprising the number and types of sampling schemes and the sample size.

With this in mind, the purpose of this paper is to provide a framework for developing sampling designs in qualitative research. In particular, we provide a typology of sampling designs for qualitative researchers. Using this typology, we introduce the following sampling strategies of inquiry: (a) parallel sampling designs, which represent a body of sampling strategies that facilitate credible comparisons of two or more different subgroups (e.g., girls vs. boys) that are extracted from the same levels of study (e.g., third-grade students); (b) nested sampling designs, which are sampling strategies that facilitate credible comparisons of two or more members of the same subgroup, wherein

one or more members of the subgroup represent a sub-sample (e.g., key informants) of the full sample; and (c) multilevel sampling designs, which represent sampling strategies that facilitate credible comparisons of two or more subgroups that are extracted from different levels of study (e.g., students vs. teachers). We show how such designs, because they facilitate comparisons, are consistent with Turner's (1980) notion that all explanation is essentially comparative and takes the form of translation of metaphors (i.e., literal translation or idiomatic translation; Barnwell, 1980). Also, we link sampling designs to various qualitative data analysis techniques (e.g., within-case analyses, cross-case analyses). We contend that our sampling framework arises from a desire to construct more adequate interpretive explanations, as well as to follow the lead of Conastas (1992), who surmised that "since we are committed to opening the private lives of participants to the public, it is ironic that our methods of data collection and analysis often remain private and unavailable for public inspection" (p. 254).

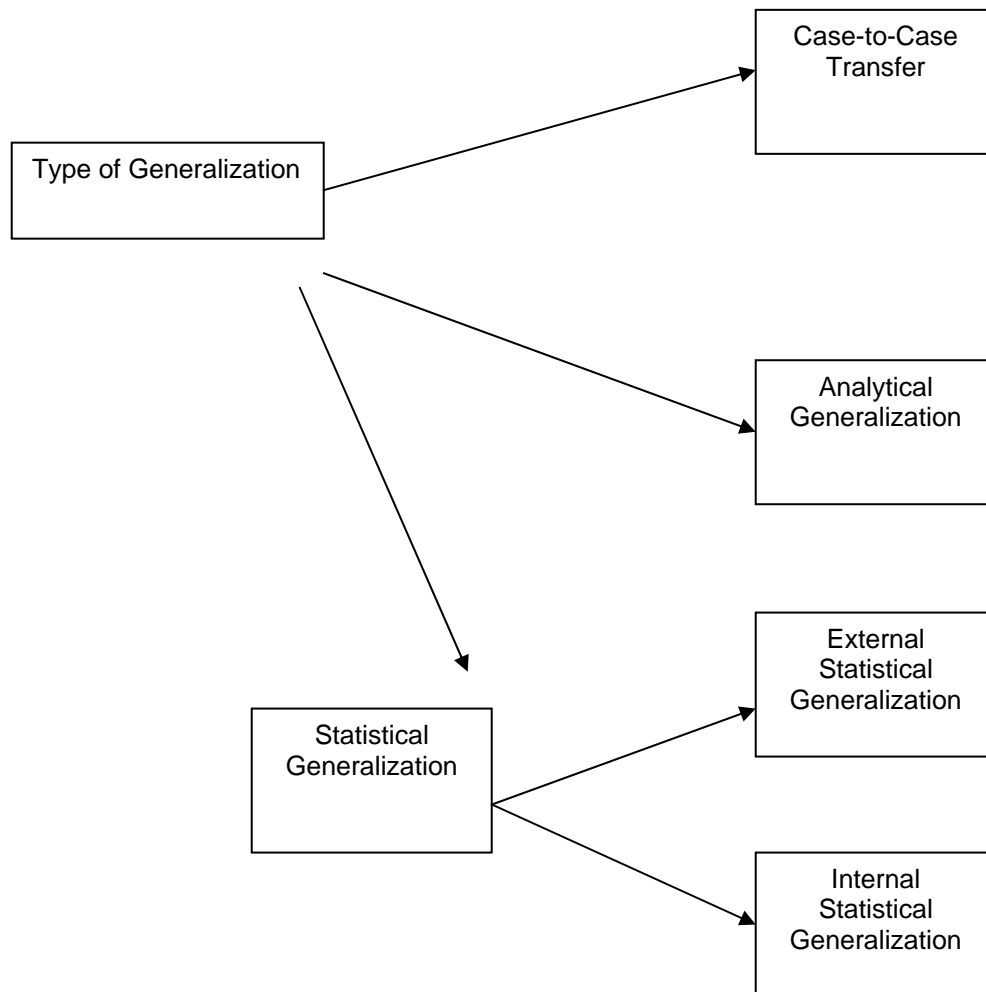
Sampling Schemes

In quantitative research, generally, only one type of statistical generalization is pertinent, namely generalizing findings from the sample to the underlying population. In contrast, in interpreting their data, qualitative researchers typically tend to make one of the following types of generalizations: (a) statistical generalizations, (b) analytic generalizations, and (c) case-to-case transfer (Curtis et al., 2000; Firestone, 1993; Kennedy, 1979; Miles & Huberman, 1994). As illustrated in Figure 1, in qualitative research, the authors believe that there are two types of statistical generalizations; external statistical generalizations and internal statistical generalizations. External statistical generalization, which is identical to the traditional notion of statistical generalization in quantitative research, involves making generalizations or inferences on data extracted from a representative statistical sample to the population from which the sample was drawn. In contrast, internal statistical generalization involves making generalizations or inferences on data extracted from one or more representative or elite participants to the sample from which the participant(s) was drawn. Analytic generalizations are "applied to wider theory on the basis of how selected cases 'fit' with general constructs" (Curtis et al., p. 1002). Finally, case-to-case transfer involves making generalizations from one case to another (similar) case (Firestone; Kennedy).

Qualitative researchers typically do not make external statistical generalizations because their goal usually is not to make inferences about the underlying population, but to attempt to obtain insights into particular educational, social, and familial processes and practices that exist within a specific location and context (Connolly, 1998). Moreover, interpretivists study phenomena in their natural settings and strive to make sense of, or to interpret, phenomena with respect to the meanings people bring (Denzin & Lincoln, 2005). However, the other three types of generalizations (i.e., internal statistical generalizations, analytic generalizations, and case-to-case transfers) are very common in qualitative research, with analytic generalizations being the most popular. More specifically, qualitative researchers "generalize words and observations... to the *population* of words/observations (i.e., the "truth space") representing the underlying context" (Onwuegbuzie, 2003, p. 400). As noted by Williamson Shafer and Serlin (2005),

The observations in any qualitative study are necessarily a subset of all other things that might have been observed using a particular set of tools and techniques in a particular setting. From this subset of all possible observations, a further subset is extracted to form the basis of qualitative inferences, since no qualitative analysis accounts for all of the observational data in equal measure. (p. 20)

Figure 1. Types of generalization in qualitative research.



Therefore, sampling is an essential step in the qualitative research process. As such, choice of sampling scheme is an important consideration that all qualitative researchers should make. Encouragingly, qualitative researchers have many sampling schemes from which to choose. Indeed, extending the work of Patton (1990) and Miles and Huberman (1994), Onwuegbuzie and Leech (2004b) identified 24 sampling schemes that are available to researchers including qualitative, quantitative, and mixed methods researchers. All of these sampling schemes can be classified as representing either random sampling (i.e., probabilistic sampling) schemes or non-random sampling (i.e.,

non-probabilistic sampling) schemes. Each of these sampling schemes is presented by sampling type (i.e., random vs. nonrandom sampling scheme) in Onwuegbuzie and Collins (2007). Although relatively rare, if the objective of the study is to generalize qualitative findings from the sample to the population, then the researcher should attempt to select a sample that is representative. Given a large enough sample, of all sampling schemes, random sampling offers the best chance for a researcher to obtain a representative sample. Thus, if external statistical generalization is the goal, which typically is not the case, then qualitative researchers should consider selecting one of the five random sampling schemes (i.e., simple random sampling, stratified random sampling, cluster random sampling, systematic random sampling, and multi-stage random sampling).

Conversely, if the goal is not to generalize to a population but to obtain insights into a phenomenon, individuals, or events, as is most often the case in interpretivist studies, then the qualitative researcher purposefully selects individuals, groups, and settings for this phase that increases understanding of phenomena. In this situation, the researcher should select one of the 19 purposive sampling schemes.

Sample Size

Even though qualitative investigations typically involve the use of small samples, choice of sample size still is an important consideration because it determines the extent to which the researcher can make each of the four types of generalizations (Onwuegbuzie & Leech, 2005b). As noted by Sandelowski (1995), “a common misconception about sampling in qualitative research is that numbers are unimportant in ensuring the adequacy of a sampling strategy” (p. 179). Nevertheless, some methodologists have provided guidelines for selecting samples in qualitative studies based on the research design (e.g., case study, ethnography, phenomenology, grounded theory) or research method (e.g., focus group). These recommendations are presented in Onwuegbuzie and Collins (2007). In general, sample sizes in qualitative research should not be too large that it is difficult to extract thick, rich data. At the same time, as noted by Sandelowski, the sample should not be too small that it is difficult to achieve data saturation (Flick, 1998; Morse, 1995), theoretical saturation (Strauss & Corbin, 1990), or informational redundancy (Lincoln & Guba, 1985).

Qualitative Sampling Designs

Most research questions in qualitative studies lead to one of two classes of analyses; within-case analyses or cross-case analyses. As delineated by Miles and Huberman (1994), within-case analyses involve analyzing, interpreting, and legitimizing data that help to explain “phenomena in a bounded context that make up a single ‘case’—whether that case is an individual in a setting, a small group, or a larger unit such as a department, organization, or community” (p. 90). In fact, within-case analyses are appropriate in samples with more than one case, providing that the researcher’s goal is not to compare the cases. As such, when a within-case analysis represents the method of choice, the researcher’s sampling design involves selection of both the sample size and sampling scheme.

On the other hand, as noted by Yin (2003), selecting multiple cases represents replication logic. That is, additional participants are chosen for study because they are expected to yield similar data or different but predictable findings (Schwandt, 2001). Stake (2000) referred to these designs as collective case studies. According to Stake, collective case studies involve the

study [of] a number of cases in order to investigate a phenomenon, population, or general condition....[who] are chosen because it is believed that understanding them will lead to better understanding, perhaps better theorizing, about a still larger collection of cases. (p. 437)

Thus, when qualitative research designs involving multiple cases are used, a major goal of the researcher is to compare and contrast the selected cases. In such instances, a cross-case analysis is a natural choice. A cross-case analysis involves analyzing data across the cases (Schwandt). Moreover, it represents a thematic analysis across cases (Creswell, 2007).

Because collective case studies typically necessitate researchers to choose their cases (Stake, 2000), being able to investigate thoroughly and understand the phenomenon of interest depends heavily on appropriate selection of each case (Patton, 1990; Stake; Vaughan, 1992; Yin, 2003). In fact, in collective case studies, “nothing is more important than making a proper selection of cases” (Stake, p. 446). Unfortunately, little or no guidance is provided in the literature as to how to select cases in collective case studies. Thus, in what follows, we introduce a typology of sampling designs that qualitative researchers might find useful when selecting participants in multiple-case studies.¹ This typology centers on the relationship of the selected cases to each other. These relationships either can be parallel, nested, or multilevel leading to parallel sampling designs, nested sampling designs, and multilevel sampling designs, respectively. Each of these classes of qualitative sampling designs is discussed in the following sections.

Parallel Sampling Designs

Parallel sampling designs represent a body of sampling strategies that facilitate credible comparisons of two or more cases. These designs can involve comparing each case to all others in the sample (i.e., pairwise sampling designs) or it can involve comparing subgroups of cases (i.e., subgroup sampling designs). Choice of these sampling designs stem from the research question(s) and the research design (e.g., case study, ethnography, phenomenology, grounded theory).

Pairwise sampling designs traditionally have been the most common types of qualitative sampling designs. These sampling designs are called “pairwise” because all the selected cases are treated as a set and their “voice” is compared to all other cases one at a time in order to understand better the underlying phenomenon, assuming that the collective voices generated by the set of cases lead to data saturation. In situations where theoretical saturation is reached, analyzing these sets of voices can lead to the generation of theory.

¹ For the purposes of this article, multiple-case studies refer to any studies that result in more than one case being selected (e.g., collective case study, ethnography, phenomenology, grounded theory).

Pairwise sampling designs can arise from any of the 24 sampling schemes. For example, the set of cases can be selected such that they represent homogeneous cases, or they can be selected to yield maximum variation. In fact, regardless of choice of sampling scheme, each case is compared to all other cases; thus, pairwise comparisons can then be undertaken.

Pairwise sampling designs lead to an array of data analysis techniques. For instance, analysts can use traditional procedures such as the method of constant comparison, keywords-in-context, word count, classical content analysis, domain analysis, taxonomic analysis, componential analysis, or discourse analysis.² In addition to these traditional analytical methods, qualitative researchers can use cross-case analytical techniques such as the following: partially ordered meta matrix, conceptually ordered displays, case-ordered descriptive meta-matrix, case-ordered effects matrix, case-ordered predictor-variable matrix, and causal networks.³

In contrast to pairwise sampling designs, subgroup sampling designs involve the comparison of different subgroups (e.g., girls vs. boys) that are extracted from the same levels of study (e.g., third-grade students). Indeed, comparing subgroups with respect to their voices is equivalent to what quantitative researchers call disaggregating data. As noted by Onwuegbuzie and Leech (2004a), comparing the voices of various subgroups in a set of cases prevents readers from incorrectly assuming that the researchers' findings are invariant across all subgroups inherent in their studies. Unfortunately, the practice of disaggregating data is underutilized by interpretivists, even though such practice is more in line with the tenet in qualitative research (than with quantitative research) of not ignoring the uniqueness and complexities of subgroups by mechanically and systematically aggregating data. Disturbingly, not examining the extent to which the voice should be disaggregated can lead to certain research subgroups being marginalized. That is, misrepresentation occurs when the commitment to generalize across the collection of cases is so dominant that the researcher's focus is unduly drawn away from aspects that are important for understanding each subgroup. Interestingly, many of the current qualitative software (e.g., NVIVO; version 7.0; QSR International Pty Ltd., 2006) make it easier for researchers to compare subgroups electronically than by hand. In fact, these software programs allow data stored in Excel files that contain demographic information to be imported for the purpose of facilitating the comparison of various subgroups. As is the case for pairwise sampling designs, when subgroup sampling is the design of choice, researchers can use traditional procedures (e.g., the method of constant comparison, componential analysis) and/or cross-case analyses (e.g., partially ordered meta matrix, case-ordered effects matrix, causal networks).

Although, technically, each subgroup can contain one case, comparing subgroups consisting of one case, when one or more of the subgroups contain an atypical case, poses a threat to what Maxwell (1996) referred to as "internal generalization" (p. 97),⁴ which

2 For reviews of traditional qualitative data analysis procedures, see Leech and Onwuegbuzie (in press) and Ryan and Bernard (2000).

3 For a review of these and other cross-case analyses, see Miles and Huberman, (1994).

4 It should be noted that what Maxwell (1996) terms "internal generalization" is not the same as what we term "internal statistical generalization." Internal generalization refers to whether conclusions drawn from the particular participants, settings, and times studied are representative of the case as a whole. In contrast internal statistical generalization denotes the extent to which the subsample members used, such as elite members and key informants, provide data that are representative of the other sample members.

refers to whether conclusions drawn from the particular participants, settings, and times examined are representative of the case as a whole. For example, if an analyst compared the voice of a typical case to that of an atypical case that was extracted using sampling schemes such as extreme case sampling and intensity sampling, then any differences extracted might not justify the researcher making internal statistical generalizations, analytical generalizations, or case-to-case transfers.⁵ Similarly, comparing subgroups containing two cases might be problematic because it might be difficult to reach information redundancy or data saturation with two cases if at least one of the cases is atypical. Therefore, we recommend that when comparing subgroups, at least three cases per subgroup should be selected. Further, the more subgroups that are compared, the larger the sample size should be. For example, a comparison of three elementary grade level subgroups (e.g., Grade 1 vs. Grade 2 vs. Grade 3) likely would necessitate a sample size of at least 9 cases (i.e., 3 subgroups x 3), whereas a comparison of four racial subgroups (e.g., African American vs. White vs. Hispanic vs. Native American) likely would call for a sample size of at least 12 cases (i.e., 3 subgroups x 4). The following six sampling schemes are best suited to subgroup sampling designs: maximum variation, homogenous sampling, critical case sampling, theory-based sampling, typical case sampling, and stratified purposeful sampling.

In addition, researchers can compare subgroups based on more than one attribute. For example, a qualitative researcher could stratify the sample by gender and by racial subgroup and then compare each gender x racial subgroup combination. For instance, four racial subgroups of interest would yield eight cells (i.e., 2 genders x 4 racial subgroups), which likely would necessitate a sample size of at least 24 participants (i.e., 8 cells x 3 cases per cell); a sample size that might be too large to obtain thick, rich description from each case, thereby preventing “the detailed reporting of social or cultural events that focuses on the ‘webs of significance’ (Geertz, 1973) evident in the lives of the people being studied” (Noblit & Hare, 1988, p. 12). This gender x racial subgroup sampling design example is displayed in Table 1.

Thus, as can be seen from Table 3, the more attributes that are used to stratify subgroups, the larger the sample size needs to be. Also, the more subgroups (within an attribute) the researcher wants to compare, the larger the sample should be. Using the table in Onwuegbuzie and Collins (2007), we suggest that researchers avoid comparing more than 4 subgroups for phenomenological studies (cf. Creswell, 1998), and more than between 7 (using Creswell’s 2002 criteria) and 10 (using Creswell’s 1998 criteria) subgroups for grounded theory studies. Also, the number of attributes stratified likely should not exceed two for phenomenological studies and five in grounded theory studies.

In order to make subgroup comparisons, cases within each subgroup should be compared to determine whether one case can be represented in terms of the other cases. That is, qualitative researchers first should examine whether meanings of one case can be reciprocally translated (i.e., literal translation or idiomatic translation; Barnwell, 1980) into the meanings of another case. As noted by Noblit and Hare (1988), translations have

5 However, it should be noted that similarities found when comparing heterogeneous cases via subgroup sampling designs could help to develop theory by facilitating a negative case analysis, which is the process of expanding and revising one’s interpretation until all outliers have been explained (Creswell, 2007; Ely, Anzul, Friedman, Garner, & Steinmetz, 1991; Lincoln & Guba, 1985; Maxwell, 1992, 1996, 2005; Miles & Huberman, 1994).

utility because they “protect the particular, respect holism, and enable comparison” (p. 28). These translations should represent, in a reduced mode, the complexity of the lived experiences of all cases that belong to a particular subgroup. In fact, we contend that making within-subgroup comparisons should be interpretive rather than aggregative, such that they will lead to the construction of adequate interpretive explanations.

Table 1

Example of Gender x Racial Subgroup Sampling Design^a

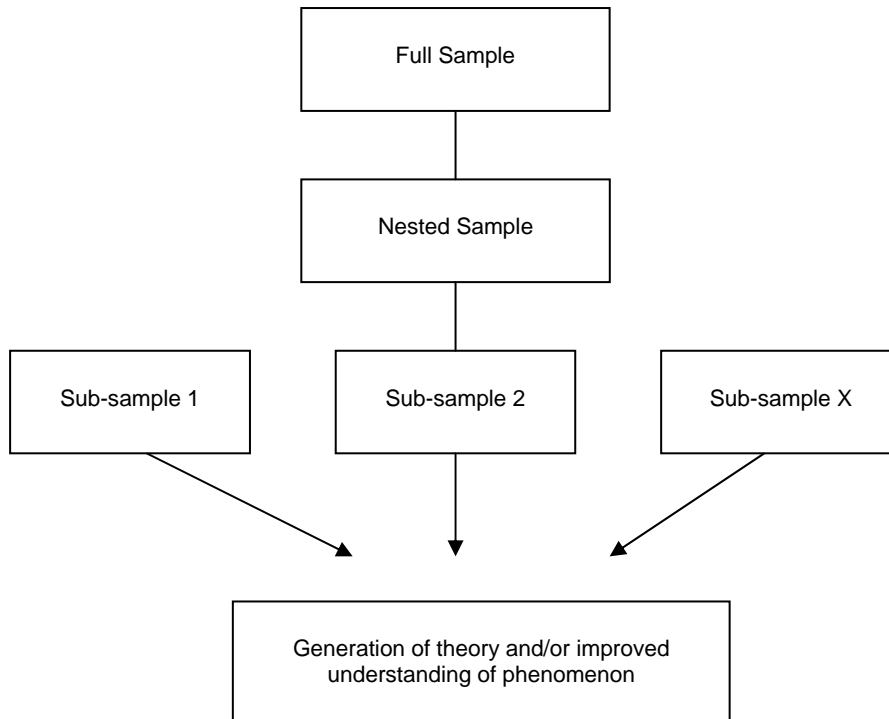
Gender	Race				Total
	African American	White	Hispanic	Native American	
Female	<i>n</i> , 3	<i>n</i> , 3	<i>n</i> , 3	<i>n</i> , 3	<i>n</i> , 12
Male	<i>n</i> , 3	<i>n</i> , 3	<i>n</i> , 3	<i>n</i> , 3	<i>n</i> , 12
Total	<i>n</i> , 6	<i>n</i> , 6	<i>n</i> , 6	<i>N</i> , 6	<i>N</i> , 24

^a This is called a 2 x 4 Subgroup Sampling Design

Nested Sampling Designs

Nested sampling designs represent sampling strategies that facilitate credible comparisons of two or more members of the same subgroup, wherein one or more members of the subgroup represent a sub-sample of the full sample. The goal of this sub-sampling is to obtain a sub-sample of cases from which further data can be extracted (cf. Figure 2). This sub-sampling often takes the form of theoretical sampling, which involves the sampling of additional people, incidents, events, activities, documents, and the like in order to develop emergent themes; to assess the adequacy, relevance, and meaningfulness of themes; to refine ideas; and to identify conceptual boundaries (Charmaz, 2000). As noted by Charmaz, “the aim of [theoretical sampling] is to refine *ideas*, not to increase the size of the original sample” (p. 519). Because theoretical sampling is the hallmark of grounded theory designs (Glaser & Strauss, 1967), nested sampling designs are particularly pertinent for grounded theorists.

Figure 2. The flow of nested sampling designs.



Nested sampling designs are most commonly used to select key informants.⁶ In fact, key informants, who are selected from the overall set of research participants, often generate a significant part of the researcher's data. Moreover, the voices of key informants often help the researcher to attain data saturation, theoretical saturation, and/or informational redundancy. Findings from key informants are generalized to the other non-informant sample members. That is, the voices of the key informants are used to make both internal statistical generalizations and analytical generalizations. The extent to which it is justified to generalize the key informants' voices to the other study participants primarily depends on how representative these voices are. Consequently, qualitative researchers must make careful decisions about their choices of key informants. Failure to make optimal sampling decisions could culminate in what some researchers refer to as key informant bias (Maxwell, 1996, 2005). Unfortunately, key informant bias is a common feature in qualitative studies because of the unrepresentativeness of key informants (Hannerz, 1992; Maxwell, 1995, 1996; Pelto & Pelto, 1975; Poggie, 1972). Some researchers recommend that key informants be selected via "systematic sampling" (Maxwell, 1996, p. 73). We believe that one way of systematically selecting key informants is by utilizing one of the four major random sampling designs (i.e., simple random sampling, stratified random sampling, cluster random sampling, systematic random sampling). Thus, we believe that on some occasions, random sampling might be appropriate for nested sampling designs. In addition to random sampling schemes, the following seven purposive sampling schemes presented by Onwuegbuzie and Collins (2007) are most appropriate for nested sampling designs: maximum variation, critical

⁶ Nested sampling designs also are useful for conducting member checks on a sub-sample of the study participants.

case sampling, theory-based sampling, typical case sampling, random purposeful sampling, multi-stage purposeful random sampling, and multi-stage purposeful sampling. Whatever sampling scheme is used to select a nested sampling design, it is important that the researcher strive to obtain representativeness via intracultural diversity (Sankoff, 1971).

Nested samples can be selected at any stage of the qualitative research process. For example, these samples can be chosen prior to the study using any of the 11 sampling schemes (i.e., four random and seven purposive). Otherwise, nested samples can be chosen during the study on one or more occasions in order to collect, to interpret, or to verify data. Therefore, nested sampling designs can be used whether the overall research is systematic (i.e., uses rigorous, pre-set procedures), emerging (i.e., allows a theory to emerge from the data instead of using specific, pre-set categories), or constructivist (i.e., focuses on the views, attitudes, beliefs, values, feelings, philosophies, and assumptions of individuals rather than concentrating on facts and describing behavior). Further, the nested samples can be chosen using attributes that were known prior to the study (e.g., demographics). Conversely, these sub-samples can be selected using data that are collected during the study. In particular, nested samples can be chosen using qualitative data or quantitative data. With respect to the former, for example, after interviewing survivors of breast cancer regarding their experiences while having treatment, the researcher could select key informants for a second round of interviewing those participants whose experiences during treatment were extreme (i.e., extreme case sampling), intense (i.e., intensity sampling), or typical (i.e., typical sampling). Alternatively, the researcher could have selected their nested samples using quantitative data. For instance, the researcher could have administered a scale that measures attitudes towards breast cancer treatment and then select several cases with the lowest and high attitude scores for the second round of interviews. When quantitative data are used to select nested designs, then the overall study changes from qualitative research to mixed methods research—specifically a sequential mixed methods design (cf. Tashakkori & Teddlie, 1998, 2003).⁷

For the same reasons as given when discussing parallel sampling designs, in general, qualitative researchers should avoid selecting sub-samples with less than three participants. In fact, assuming the goal is for the key informants to be representative of the study participants as a whole, the larger the sample, the larger the nested sample should be. Also, as is the case for parallel sampling designs, nested samples can be selected using more than one attribute. When comparing nested samples, researchers can use traditional procedures and/or cross-case analyses. Regardless of the analyses used, in making such comparisons, the researcher should examine whether meanings of one case can be reciprocally translated into other cases.

Multilevel Sampling Designs

Finally, multilevel sampling designs represent sampling strategies that facilitate credible comparisons of two or more subgroups that are extracted from different levels of study. For example, a qualitative researcher might be interested in comparing the

⁷ Onwuegbuzie and Teddlie (2003) refer to designs that use quantitative data to generate qualitative data as sequential (mixed) quantitative-qualitative analyses—specifically, quantitative extreme case analyses.

perceptions of students regarding standardized tests to those of their teacher(s). Clearly, the student and teacher samples represent some form of hierarchy. Because of this hierarchy, the sampling schemes and sample sizes used for the lower-level and upper-level samples/sub-samples typically are not uniform. For example, because students represent the lower-level sample/sub-sample and their teacher(s) represents the upper-level, it is not uncommon for the voices of several students to be compared with the voice of one teacher. Further, whereas the student participants (i.e., lower level sample/sub-sample) might be selected using any of the 24 sampling schemes, the teacher likely would be selected either via convenient sampling (e.g., if the students' teacher(s) is selected), critical case sampling, politically important case sampling, or criterion sampling, or by using one of the four random sampling techniques, in situations where the researcher has a pool of teachers from which to select the upper level sample/sub-sample.

Although it is possible for qualitative researchers to select lower-level and upper level samples/sub-samples that are independent (e.g., students from one school and teacher(s) from another school in the same or another school district), the lower-level and upper level samples/sub-samples usually are very much related to each other. Moreover, these samples/sub-samples tend to be conditionally related. That is, once one level is selected (e.g., students), then the other is automatically selected (e.g., students' teacher(s)). However, whatever the relationship is between the multilevel samples, when comparing the samples/sub-samples, the researcher should examine whether meaning extracted from one sample/sub-sample can be reciprocally translated into the meanings of the other sample(s)/sub-sample(s). Because of the hierarchical structure of the samples/sub-samples, hierarchical qualitative analyses could be considered, such as those described by Onwuegbuzie (2003) that include the extraction of "meta-themes," which represent themes at a higher level of abstraction than the original emergent themes.

Conclusions

In comparing cases, whether via parallel sampling designs, nested sampling designs, or multilevel sampling designs, it is essential that qualitative researchers do not sacrifice "thick description" (Geertz, 1973) for comparative description by focusing only on comparing a limited number of attributes, feelings, experiences, thoughts, opinions, events, activities, experiences, and/or processes. In other words, in an effort to make comparisons, the uniqueness and complexity of each case should not be trivialized (Stake, 2000). Moreover, these comparisons should be emergent, interactive, and flexible. Regardless of the sampling design, qualitative researchers should not select their cases merely for purpose of comparison. Rather, each case should be chosen because of her/his own intrinsic and unique value (Stake). Thus, in qualitative studies involving multiple cases, qualitative researchers must strike a fine balance between obtaining thick description from each case and obtaining comparative description from each comparison.

Qualitative researchers contend that context affects the meaning of events. Because context often varies for different subgroups, comparing subgroups is a technique that has the potential of helping researchers to maximize their understanding of phenomena. Therefore, this paper has provided a framework for making comparisons in qualitative research by introducing a typology of sampling designs for qualitative researchers. Inherent in this framework is a typology of sampling schemes and guidelines

for selecting sample sizes. We believe that the greatest appeal of this framework is that it can help researchers to identify an optimal sampling design for their qualitative studies. In particular, our framework can be used to inform sampling decisions made by the researcher such as selecting a sampling scheme (e.g., random vs. purposive), selecting an appropriate sample size, and selecting an appropriate sampling strategy (i.e., parallel, nested, and/or multilevel) that enable appropriate generalizations (i.e., external statistical, internal statistical, analytical, face-to-face transfer) to be made relative to the study's design. Using our framework also provides a means for making these decisions explicit and promotes interpretive consistency between the interpretations made in qualitative studies and the sampling design used, as well as the other components that characterize the formulation (i.e., goal, objective, purpose, rationale, and research question), planning (i.e., research design), and implementation (i.e., data collection, analysis, legitimation, and interpretation) stages of the qualitative research process. Thus, we hope that researchers from the social and behavioral sciences and beyond consider using our sampling design framework so that they can design qualitative inquiries in ways that address the challenges of representation, legitimation, and praxis.

Some interpretivists might reject our typology because they believe that comparing subgroups represents a shift to the quantitative paradigm (e.g., analysis of variance). However, rather than representing a paradigm shift, we contend that our typology represents an elaboration of existing understanding about sampling cases and analyzing data. Thus, we hope that other qualitative research methodologists build on the typology provided in this article and/or develop other typologies that help bring qualitative researchers closer to "verstehen".

References

- Anfara, V. A., Brown, K. M., & Mangione, T. L. (2002). Qualitative analysis on stage: Making the research process more public. *Educational Researcher*, 31(7), 28-38.
- Barnwell, K. (1980). *Introduction to semantics and translation*. Horleys Green, England: Summer Institute of Linguistics.
- Baumgartner, T. A., Strong, C. H., & Hensley, L. D. (2002). *Conducting and reading research in health and human performance* (3rd ed.). New York: McGraw-Hill.
- Bernard, H. R. (1995). *Research methods in anthropology: Qualitative and quantitative approaches*. Walnut Creek, CA: AltaMira.
- Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509-535). Thousand Oaks, CA: Sage.
- Collins, K. M. T., Onwuegbuzie, A. J., & Jiao, Q. G. (2006). Prevalence of mixed methods sampling designs in social science research. *Evaluation and Research in Education*, 19, 83-101.
- Collins, K. M. T., Onwuegbuzie, A. J., & Jiao, Q. G. (2007). A mixed methods investigation of mixed methods sampling designs in social and health science research. *Journal of Mixed Methods Research*, 1, 267-294.
- Connolly, P. (1998). "Dancing to the wrong tune": Ethnography generalization and research on racism in schools. In P. Connolly & B. Troyna (Eds.), *Researching*

- racism in education: Politics, theory, and practice* (pp. 122-139). Buckingham, UK: Open University Press.
- Constas, M. A. (1992). Qualitative data analysis as a public event: The documentation of category development procedures. *American Educational Research Journal*, 29, 253-266.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson Education.
- Creswell, J. W. (2007). *Qualitative inquiry and research method: Choosing among five approaches* (2nd. ed.). Thousand Oaks, CA: Sage.
- Crowley, E. P. (1994) Using qualitative methods in special education research. *Exceptionality*, 5, 55-70.
- Curtis, S., Gesler, W., Smith, G., & Washburn, S. (2000). Approaches to sampling and case selection in qualitative research: Examples in the geography of health. *Social Science and Medicine*, 50, 1001-1014.
- Denzin, N. K. (1978). *Sociological methods: A sourcebook*. New York: McGraw-Hill.
- Denzin, N. K., & Lincoln, Y. S. (2005). The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (3rd ed., pp. 1-32). Thousand Oaks, CA: Sage.
- Ely, M., Anzul, M., Friedman, T., Garner, D., & Steinmetz, A. C. (1991). *Doing qualitative research: Circles within circles*. New York: Falmer.
- Firestone, W. A. (1993). Alternative arguments for generalizing from data, as applied to qualitative research. *Educational Researcher*, 22(4), 16-23.
- Flick, U. (1998). *An introduction to qualitative research: Theory, method, and applications*. London: Sage.
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In C. Geertz (Ed.), *The interpretation of cultures: Selected essays* (pp. 37-126). New York: Basic Books.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Guba, E. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology*, 29(2), 75-91.
- Hannerz, U. (1992). *Cultural complexity: Studies in social organization of meaning*. New York: Columbia University Press.
- Johnson, R. B., & Christensen, L. B. (2004). *Educational research: Quantitative, qualitative, and mixed approaches*. Boston, MA: Allyn and Bacon.
- Jones, S. R. (2002). Writing the word: Methodological strategies and issues in qualitative research. *Journal of College Student Development*, 43, 461-473.
- Kennedy, M. (1979). Generalizing from single case studies. *Evaluation Quarterly*, 3, 661-678.
- Krueger, R. A. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). Thousand Oaks, CA: Sage.
- Langford, B. E., Schoenfeld, G., & Izzo, G. (2002). Nominal grouping sessions vs. focus groups. *Qualitative Market Research*, 5, 58-70.

- Leech, N. L., & Onwuegbuzie, A. J. (in press). An array of qualitative data analysis tools: A call for qualitative data analysis triangulation. *School Psychology Quarterly*.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62, 279-299.
- Maxwell, J. A. (1995, February). *Diversity and methodology in a changing world*. Paper presented at the Fourth Puerto Rican Congress of Research in Education, San Juan, Puerto Rico.
- Maxwell, J. A. (1996). *Qualitative research design*. Newbury Park, CA: Sage.
- Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Newbury Park, CA: Sage.
- Merriam, S. B. (1995). What can you tell from an N of 1?: Issues of validity and reliability in qualitative research. *PAACE Journal of Lifelong Learning*, 4, 51-60.
- Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.): Vol. 16. Qualitative research methods series. Thousand Oaks, CA: Sage.
- Morse, J. M. (1994). Designing funded qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 220-235). Thousand Oaks, CA: Sage.
- Morse, J. M. (1995). The significance of saturation. *Qualitative Health Research*, 5, 147-149.
- Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies*: Vol. 11. Qualitative research methods series. Newbury Park, CA: Sage.
- Onwuegbuzie, A. J. (2003). Effect sizes in qualitative research: A prolegomenon. *Quality & Quantity: International Journal of Methodology*, 37, 393-409.
- Onwuegbuzie, A. J., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, 12(2), 281-316. Retrieved August 31, 2007, from <http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf>
- Onwuegbuzie, A. J., & Leech, N. L. (2004a). Enhancing the interpretation of "significant" findings: The role of mixed methods research. *The Qualitative Report*, 9, 770-792. Retrieved March 8, 2005, from <http://www.nova.edu/ssss/QR/QR9-4/onwuegbuzie.pdf>
- Onwuegbuzie, A. J., & Leech, N. L. (2004b, February). *A call for qualitative power analyses*. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX.
- Onwuegbuzie, A. J., & Leech, N. L. (2005a). Taking the "Q" out of research: Teaching research methodology courses without the divide between quantitative and qualitative paradigms. *Quality & Quantity: International Journal of Methodology*, 39, 267-296.
- Onwuegbuzie, A. J., & Leech, N. L. (2005b). The role of sampling in qualitative research. *Academic Exchange Quarterly*, 9, 280-284.
- Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed*

- methods in social and behavioral research* (pp. 351-383). Thousand Oaks, CA: Sage.
- Patton, M. Q. (1990). *Qualitative research and evaluation methods* (2nd ed.). Newbury Park, CA: Sage.
- Pelto, P., & Pelto, G. (1975). Intracultural diversity: Some theoretical issues. *American Ethnologist*, 2, 1-18.
- Poggie, J. J., Jr. (1972). Toward control in key informant data. *Human Organizational*, 31, 23-30.
- QSR International Pty Ltd. (2006). *NVIVO: Version 7. Reference guide*. Doncaster Victoria: Australia: Author.
- Ryan, G. W., & Bernard, H. R. (2000). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 769-802). Thousand Oaks, CA: Sage.
- Sandelowski, M. (1995). Focus on qualitative methods: Sample sizes in qualitative research. *Research in Nursing & Health*, 18, 179-183.
- Sankoff, G. (1971). Quantitative aspects of sharing and variability in a cognitive model. *Ethnology*, 10, 389-408.
- Schwandt, T. A. (2001). *Dictionary of qualitative inquiry* (2nd ed.). Thousand Oaks, CA: Sage.
- Stake, R. E. (2000). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 435-454). Thousand Oaks, CA: Sage.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*: Vol. 46. Applied social research methods series. Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.
- Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research*, 1, 77-100.
- Turner, S. (1980). *Sociological explanations as translation*. New York: Cambridge University Press.
- Vaughan, D. (1992). Theory elaboration: The heuristics of case analysis. In C. C. Ragin & H. S. Becker (Eds.), *What is a case? Exploring the foundations of social inquiry* (pp. 173-292). Cambridge, U.K.: Cambridge University.
- Williamson Shafer, D., & Serlin, R. C. (2005). What good are statistics that don't generalize? *Educational Researcher*, 33(9), 14-25.
- Yin, R. K. (2003). *Case study research: Design and methods*: Vol. 5. Applied social research methods series. Thousand Oaks, CA: Sage.

Author Note

Anthony Onwuegbuzie, Ph.D., is professor in the Department of Educational Leadership and Counseling at Sam Houston State University. He teaches courses in

doctoral-level qualitative research, quantitative research, and mixed methods. His research topics primarily involve disadvantaged and under-served populations such as minorities, children living in war zones, students with special needs, and juvenile delinquents. Also, he writes extensively on qualitative, quantitative, and mixed methodological topics.

Dr. Nancy L. Leech is an assistant professor at the University of Colorado at Denver and Health Sciences Center. Dr. Leech is currently teaching master's- and Ph.D.-level courses in research, statistics, and measurement. Her areas of research include promoting new developments and better understandings in applied methodology.

Correspondence should be addressed to Anthony J. Onwuegbuzie, Department of Educational Leadership and Counseling, Box 2119, Sam Houston State University, Huntsville, Texas, 77341-2119; Email: tonyonwuegbuzie@aol.com

Copyright 2007: Anthony Onwuegbuzie, Nancy L. Leech, and Nova Southeastern University

Article Citation

Onwuegbuzie, A., & Leech, N. L. (2007). Sampling designs in qualitative research: Making the sampling process more public. *The Qualitative Report*, 12(2), 238-254. Retrieved [Insert date], from <http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie1.pdf>
