2.1

Sampling

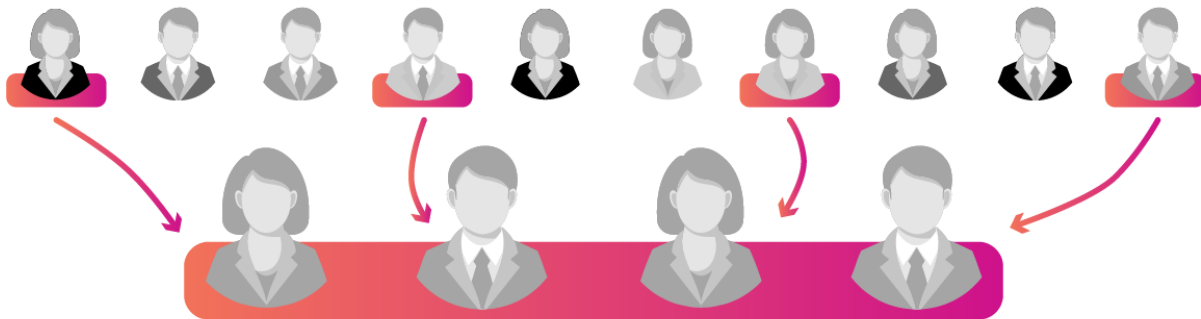### **How should I choose a sample for my study?**

Royce Kimmons

Most research studies in education require some form of sampling. Because you can't always study everyone or everything, sampling means that you only study part of a larger group and (hopefully) are still able to draw meaningful conclusions.

For instance, if you wanted to study university students over the age of 50, you might randomly survey a few hundred people by phone and then draw conclusions about other people within that demographic group more generally. In such a study, you would be concerned about two different groups of people: the larger population you are trying to draw conclusions about (designated by a capital *N*) and the smaller sample you are actually studying (designated by a lower-case *n*).

How we go about sampling is very important and will be dependent upon a variety of factors, such as the questions we're trying to answer, what we're studying, and the context. Because goals and needs vary, there is not a single, correct way to sample, but there are appropriate (and inappropriate) ways to sample depending on our study.

In this chapter, I'll explain four basic types of sampling — convenience, purposeful, random, and census — along with guidance on specific approaches for each type, as well as examples and cautions. I will then conclude by providing some considerations that should guide you in selecting appropriate approaches in your own research.

Key Terms

Census
> A study that includes every member of the population (removing the need for sampling, generalizability, etc.).

Convenience Sampling
> An approach to sampling, common in design and professional practices, that chooses subjects that are accessible to the researcher, such as testing with a colleague, interviewing a spouse, etc.

Population
> The group facing the problem researchers are trying to solve or to whom researchers will generalize their results (e.g., K-12 students), represented by the uppercase "N" variable.

Purposeful Sampling
> An approach to sampling, common in qualitative research, that chooses subjects that will provide insight for answering the study questions.
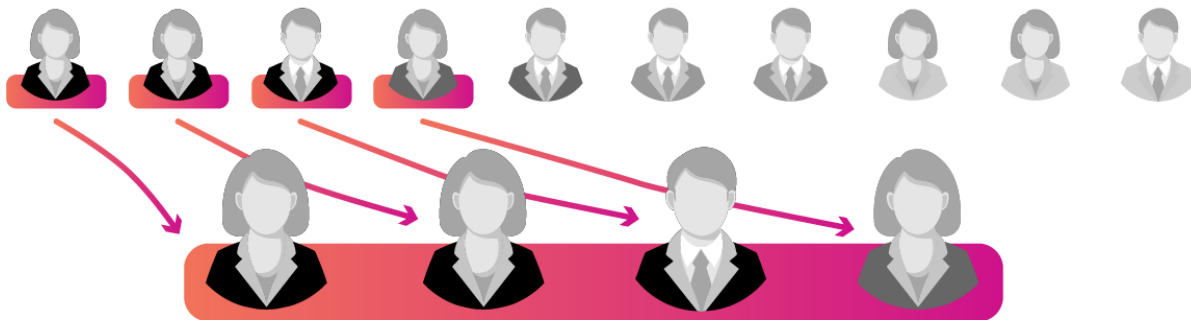
Random Sampling
> An approach to sampling, common in quantitative research, that chooses subjects randomly from a target group or population.

Sample
> The actual group being studied, represented by the lowercase "n" variable.

Convenience Sampling

Convenience sampling consists of studying those who are close to us or who are easy to study. A web developer might put an in-progress website in front of other web developers for feedback, a graduate student might survey other students in a class to get their opinions on educational reform, or a teacher might ask other teachers about their ideas for lesson improvement. In all of these cases, the primary factor guiding who is sampled is the ease at which data can be collected. Because it makes data collection easy, convenience sampling is common in user testing, design settings, and other situations that utilize non-traditional or less-formal research methodologies, such as action, design, or guerrilla research.

Of all the sampling methods, convenience sampling tends to be the easiest to do, which is why it is commonly used in teaching and design scenarios, but it is also the least accurate and most

problematic. It is great for getting quick data and guidance on topics that might deal with universal human experiences, but its results are not generalizable to the population and also do not allow you to acknowledge varieties in the population.

In the web developer example above, putting an educational site in front of a few colleagues might provide you with useful, instant guidance on how to improve some aspects of the design, but because your sample subjects will tend to be like you, it will often fail to provide you with guidance to meet the needs of all potential users (e.g., you will not know how a person with a vision disability might struggle with the site unless one of your close colleagues has a vision disability). Thus, results derived from convenience sampling will typically not meet rigor expectations of formal research, but they can be helpful for the initial shaping of research questions as well as iterative designs of instruments, tools, and products.

Purposeful Sampling

Purposeful sampling consists of selecting subjects specifically out of a population that the researcher believes will help them to most meaningfully and accurately answer the research question. A hallmark of qualitative methodologies, purposeful sampling takes at least four common forms, which I will now explain: informant, extreme case and intensity, quota, and snowball (see Table 1).

Table 1

Four Common Non-Random-Sampling Strategies with Examples

Example numbers are used for illustration only and do not accurately reflect necessary comparative sample sizes needed for each approach as this would be dependent upon the methodology, research question, and other considerations. For appropriate sample sizes, consult other studies that use similar methodologies to your own.

| Approach	Strategy Example	Sample
Informant	Interview students representing minoritized ethnic groups.	10 students from the target ethnic group
Extreme Case and Intensity	Interview students representing both majority and minoritized groups.	5 students from the minoritized group + 5 students from the majority group = 10 students
Quota	Interview the same number of students from each target ethnic group.	2 students x 5 different ethnic groups = 10 students
Snowball	Interview a student representing a target ethnic group, then ask them to help you to identify additional students in the target group, and so forth.	1 student from the target group + 2 of that student's friends + 7 of their friends = 10 students

Informant
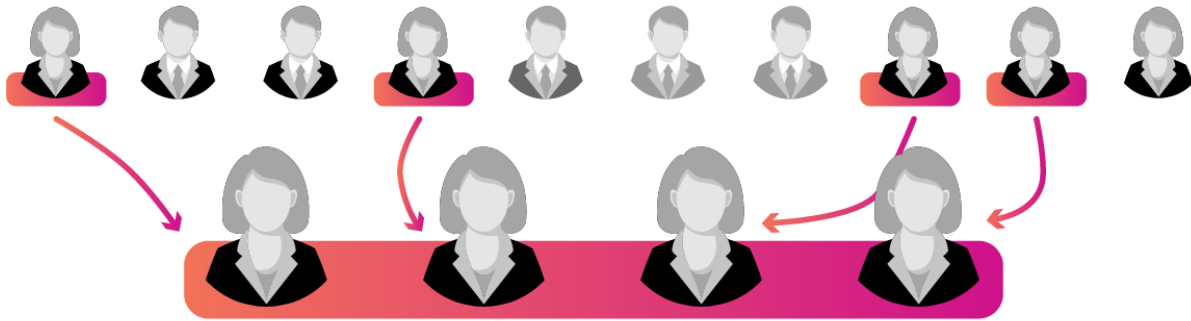
Informant sampling selects subjects based upon their expertise or ability to provide insight for answering the research question. They are intentionally selected from the population for their experience with the topic under study and their ability to help the researcher understand it. For example, if we wanted to understand women's experiences of being harassed in academia, we might interview women who have had experiences of being harassed. Doing so would not allow us to generalize their results to all women or to say how prevalent harassment is in academia, but it would allow us to understand the nuances of harassment, what it means to be harassed, what the effects of harassment might be, and how to identify it in practice.

The strength of informant sampling is that it allows us to develop a deep understanding of our topic from those best equipped to help us understand it. Yet, any results we derive from such an approach will not be reliably attributable to the whole population and will be dependent upon the criteria we use for selecting our informants. If in the case above we only select ethnic majority women to study, for instance, then we will not understand how ethnic or racial differences among women influence the forms that harassment takes, and we also will not understand whether and how men might face harassment in similar contexts. This is not a weakness of the sampling approach so much as it is an intentional delimitation, allowing the researcher to hone in on the issue or sub-population they desire to study the most.

Extreme Case and Intensity
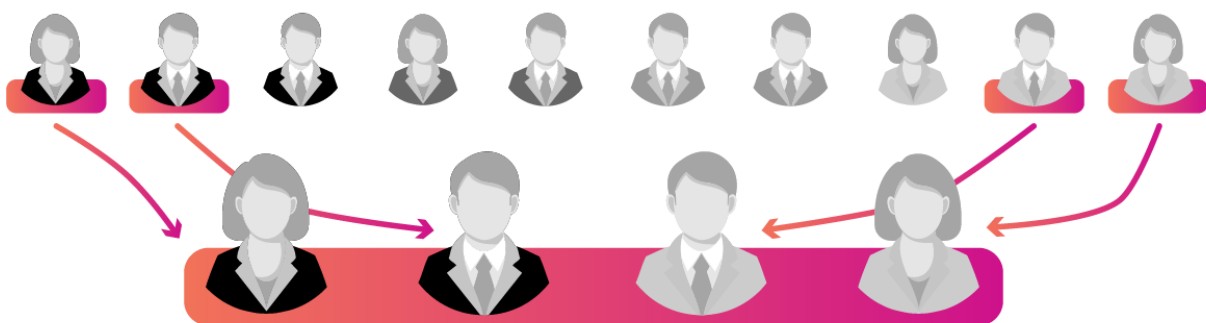
Extreme case sampling selects subjects from opposite ends or contradictory sides of the phenomenon being studied to give a sense for the breadth of the topic and divergence in experiences, opinions, or characteristics. This requires the researcher to have a sense for what the extremities might be prior to selection but allows them to show a range in their sampling. For example, if you wanted to study

how political affiliations of teachers influence their pedagogy, then you might sample teachers from both the far-right and the far-left of the political spectrum. This would allow you to effectively see how drastically experiences and effects diverge but would give you limited insight into more moderate cases of political divergence, which would likely represent the majority of the population.

As a more moderate approach, intensity sampling also tries to account for a diversity of subjects in the sample, but it is less interested in the extremity of the cases and is more focused on identifying richness in sampled subjects. To apply intensity sampling to the previous example, we might still select subjects representing the major political parties but will focus on those who will help us to best understand the phenomenon in question. For instance, interviewing an extreme case of a politically anarchistic mathematics teacher might not be very helpful for understanding the phenomenon in question, because there might not be much overlap in the teacher's political stances and how they teach fractions. However, interviewing intensity samples of politically active social studies teachers would likely be more informative even though subjects might be more politically moderate simply because their subject areas would allow the researcher to explore a richer relationship between each teacher's politics and how they teach their related subject area.

Quota
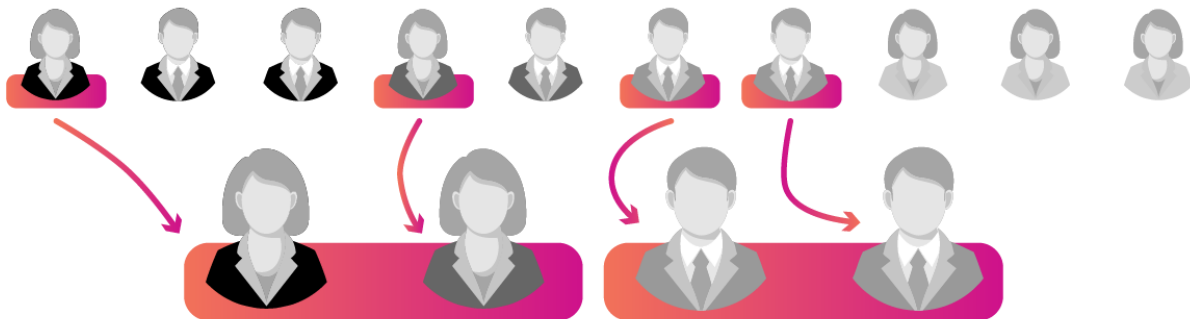
Quota sampling selects fixed or equivalent numbers of subjects from predetermined groups to ensure that the study accounts for important differences between the groups. This requires the researcher to make *a priori* assumptions about what types of differences between subjects are important, and the researcher then will treat each group of subjects like a separate population pool. For example, if you wanted to understand teacher attitudes toward hiring practices and to see how experiences might differ based on the teacher's gender, you might choose to interview five recently-hired women and five recently-hired men to see commonalities and differences between their accounts. Similarly, if you were studying the experiences of racially minoritized students in low-SES schools, you might choose to interview four students from each major racial group represented in your school's demographic data (e.g., four Latinx students, four Black students, and four Asian American students). In such approaches to sampling, the overall representation of the subgroup in the larger population is irrelevant (e.g., the Asian American student population of the school might be much smaller than the Black student population), but the goal would be to make sure that you have sampled sufficiently from each group to be able to draw meaningful conclusions.

Quota sampling is useful if researchers have *a priori* categories of analysis that matter to their research questions and if researchers are able to identify subjects based on these categories. By

conducting parallel analyses of groups of subjects, researchers can construct group-based narratives for comparison and ensure that diverse perspectives (in terms of the grouping criterion) are explored in a focused manner. If, for instance, in the hiring example above, researchers only sampled one woman and nine men, then women's hiring experiences likely would not be well-represented in the findings, but by setting a quota of five women, researchers can help ensure that women's experiences will be treated equally to men's.

If, however, our *a priori* categories are messy, poorly defined, or unnecessary to our research questions, then quota sampling is less useful. In the example of racially minoritized students, for instance, sampling four Asian American students might not yield very helpful results once we realize that the experiences of students from Korea, China, Japan, and various Pacific islands might be quite distinct or that the experiences of Black students from poor families might be very different from those of Black students from more affluent families. Thus, the key to quota sampling is to ensure that your *a priori* categories are important, accurate, and clear.

Snowball
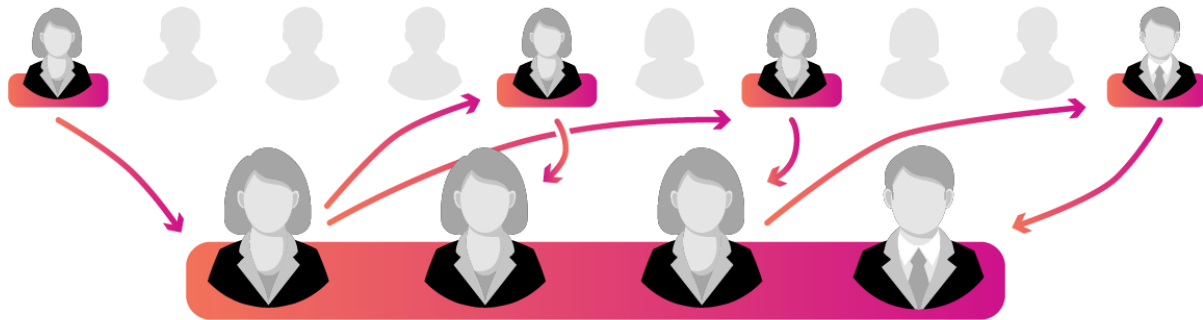
Snowball sampling selects a small number of initial subjects and then gradually grows the sample via additional sampling through the subjects. This approach is especially useful if researchers have difficulty identifying or recruiting subjects in the target population (due to ignorance, lack of trust, or outsider status). For instance, if you were studying homeschool parents, you might not be able to identify many subjects at the outset of the study due to lack of knowledge, privacy, etc. So, you might begin by only identifying one parent, interviewing them, and then, as part of your interview protocol, asking them if they know of other parents who homeschool as well.

This approach allows the researcher to utilize subjects' knowledge and social networks as sampling tools and is very useful for gaining entry to insular, invisible, or poorly understood populations. However, because it relies so heavily upon the knowledge and networks of individual subjects, it may have difficulty providing representation of the whole population. Following the homeschoolers example above, the subsequent parents that we sample will be dependent upon the initial parent, which would mean that the researchers might only ever be able to identify subjects that are like the original. If there are different networks of homeschool parents (e.g., religious homeschool communities vs. military families), then this sampling strategy might only allow entry to one of the networks and not others, and if the networks exhibit different values, behaviors, or norms from one another, then results will not account for the totality of homeschool parent experiences. Thus, as researchers traverse these networks, they should be aware of the potential for insularity in selection

and attempt to identify alternative networks that might also exist in the population.

Random Sampling

Random sampling consists of selecting a sufficient number of subjects at random from a population to ensure that results are indicative of the population at large. If your research goal is to generalize your results from your sample to an entire population, then you must utilize a random sampling strategy. Failure to randomize your sample would mean that your study might be biased toward particular sub-groups in the population (e.g., those that are near you or are like you).

For instance, if I wanted to know how students feel about their school lunch programs, the easiest thing to do would be to interview my own kids or my neighbors' kids. But would this tell me anything about school lunch programs more broadly?

Similarly, if I wanted to understand teacher beliefs about technology integration in my state, and I proposed to survey teachers via email, which teachers I sent the survey to (and which teachers responded) would be very important. If I only sent the survey to teachers in one subject area, geographic location, grade level, school socioeconomic classification, etc., then my results would not reflect the beliefs of teachers in my state broadly. And even if I sent the survey to every teacher in the state and only 10% responded, might there be reasons that the other 90% did not respond that would influence their beliefs about technology integration (such as not having a computer to complete the survey with)?

The goal of randomization is to allow you to study a relatively small proportion of the population and get results that can reasonably be said to represent the population more broadly. This makes research studies possible that would otherwise be infeasible due to exorbitant cost or other limitations (e.g., there's no way that you could ever interview every child about how they feel about school lunches), but for randomization to work for achieving generalization, it must be done in a manner that is appropriate for your research question, context, and population.

Education researchers employ at least four common random-sampling strategies for doing this well. These include simple, stratified, proportional, and cluster sampling (see Table 2), which I will now explain in more detail.

Table 2

Four Common Random-Sampling Strategies with Examples

Example numbers are used for illustration only and do not accurately reflect necessary comparative sample sizes needed for each approach as this would be dependent upon population sizes and between- and within-group variances.

Approach	Strategy Example	Sample
Simple	Survey random teachers from a state-level list of teachers.	1,000 teachers
Stratified	Survey teachers from each type of school.	334 teachers from traditional public schools 333 teachers from private schools 333 teachers from charter schools = 1,000 teachers
Proportional	Survey a representative number of teachers from each type of school based on the teacher population of each type of school.	800 teachers from traditional public schools 100 teachers from private schools 100 teachers from charter schools = 1,000 teachers
Cluster	Survey all teachers from randomly-selected schools.	300 teachers from school A 300 teachers from school B 400 teachers from school C = 1,000 teachers

Simple
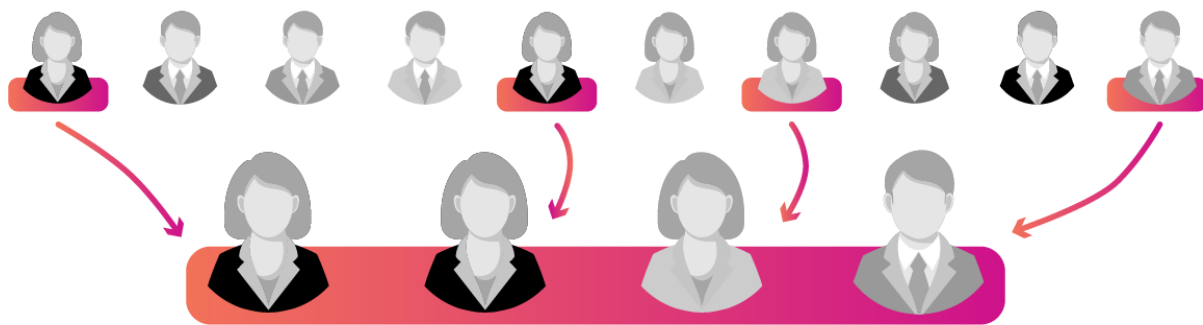
Simple random sampling selects subjects from the target population without consideration for sub-groups, categories, or differences between the subjects. This allows researchers to show that their results are representative of the population broadly while giving them limited power to answer more nuanced questions about sub-groups.

In the teacher email survey example, you might acquire a list of teacher emails from your state department of education and email a random selection of 1,000 teachers from the list. This would give you a sense for how teachers in your state broadly think about technology integration, but suppose that you wanted to then see if teacher beliefs varied based on whether they worked for a charter vs. a traditional public school. If your random sample only provided results from 10 teachers representing charter schools, then you probably would not have enough data to meaningfully compare their results to the group at large.

As this case illustrates, simple randomization is great for fast generalizability, but it will often not provide sufficient data for more nuanced analysis especially if our questions deal with sub-groups that are relatively small or that might be less likely to show up in the data. Charter schools represent a small proportion of schools generally, and so results for their teachers will not be well-represented in the dataset. Similarly, results from teachers representing racial or ethnic minoritized groups would also likely be limited for comparison, because such a high proportion of the overall sample will

represent the majority group. And teachers with limited access to technology would likely be ignored altogether because the sampling procedure did not take into account the role that email access might play in ensuring that those teachers' responses were actually counted.

If such considerations are not pertinent to the research question and context, however, then simple randomization might be appropriate and tends to be the easiest method for collecting generalizable data quickly and efficiently.
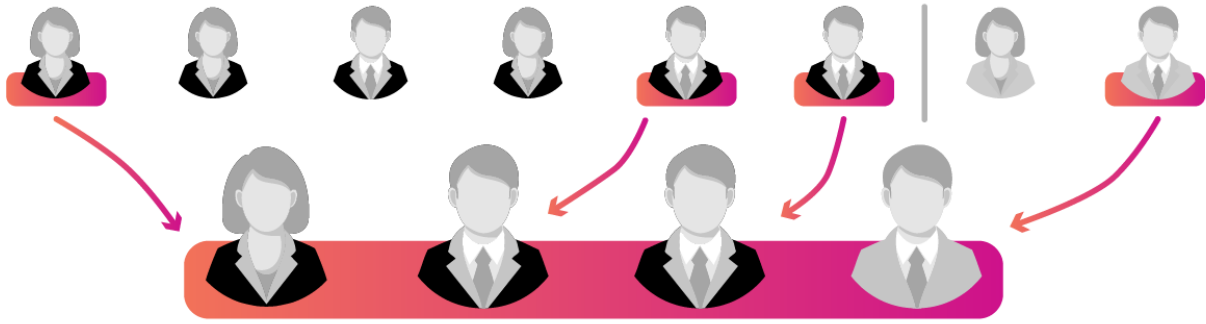
Stratified
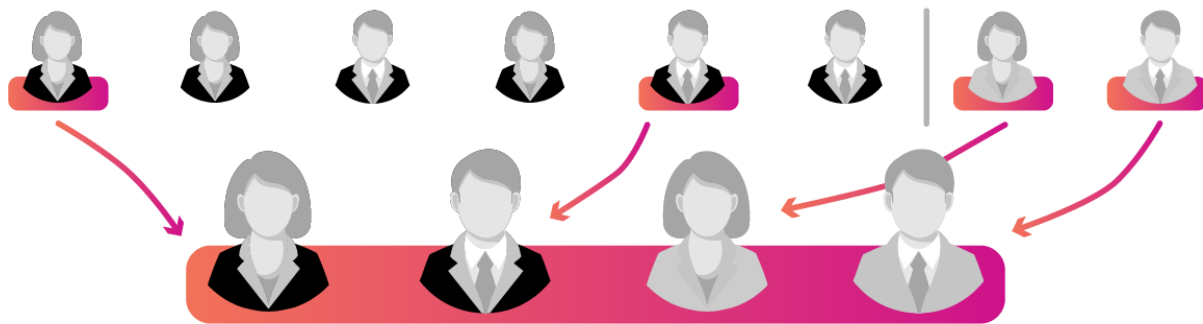
Stratified sampling selects subjects from different groups, levels, or strata in the population so that results will account for important variation in the population. Like quota sampling above, this requires *a priori* identification of the strata that differentiate subjects from one another and for researchers to then sample an appropriate number of subjects from each stratum. This allows for the sample to reflect the diversity of the larger population and prevents findings for smaller groups in the population to be ignored due to lack of representation.

For instance, if you wanted to study high school drop-out rates and believed that rates might differ based on the type of high school (e.g., charter vs. private vs. general public), then doing a simple random sample would probably result in general public high schools comprising most of your dataset. If charter schools only represented 1% of the schools in your state, then they would only be expected to represent 1% of your dataset, which might make it difficult to draw robust statistical conclusions about charter schools. To combat this, you might intentionally sample in such a way that 33.3% of your dataset represented each type of school. This would allow you to compare results between groups.

Stratified sampling is preferable to simple random sampling when you want to be able to compare results between *a priori* groups. To do it properly, though, you need to ensure that your sample size for each stratum is sufficient to make generalizable claims, and doing so can be a double-edged sword in that it allows you to draw conclusions about relatively small segments of the population, but you will then need to be careful of how you interpret results, because stratum-based results (especially for strata with smaller populations) may not generalize to the entire population.

Proportional

Like stratified sampling, probability-proportional-to-size (or PPS) sampling (also known as proportional stratified random sampling) selects subjects from different groups, levels, or strata in the population but does so at a rate proportional to their overall representation in the population. This allows for the sample to reflect key characteristics of the population at large that could have been lost through simple random sampling and also allows results to be comparable between strata and also generalizable to the entire population.

For instance, if you wanted to survey teachers on their experiences with home-life balance, you could reasonably assume that gender would be an important factor that would influence results, due to different social expectations that are placed on women vs. men. This means that if the representation of women vs. men in your sample did not match that of your population, then your results would be less accurate. Similarly, if you wanted to study student achievement on a standardized test across a state, you could reasonably assume that results would be influenced by the type of school students attend (traditional public vs. private vs. public charter). So, if your state's student population has a breakdown of 80%/10%/10% for public/private/charter, then you would want to ensure that your sample has a similar breakdown to be valid. This would mean sampling from all three types of schools until you could be confident that the sample accurately reflects the population at large.

PPS sampling is superior to simple random sampling when key *a priori* strata are identifiable at the outset because it can help to ease sampling errors that are always possible with randomization (e.g., accidentally overrepresenting men or private schools in a dataset).

Cluster
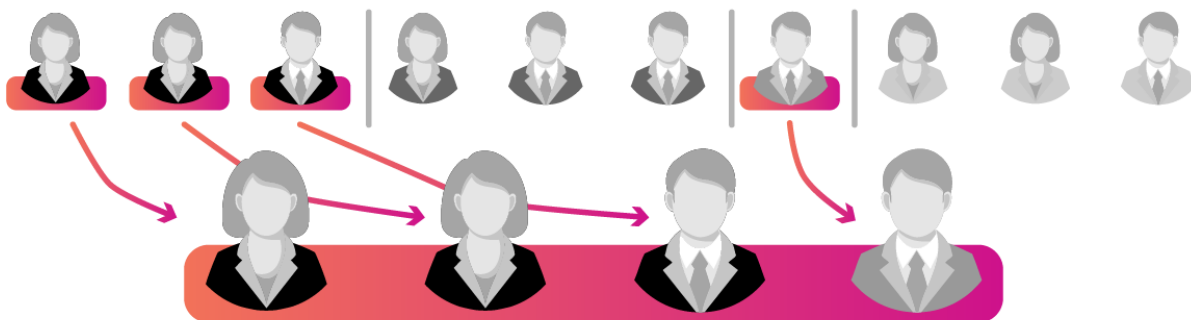
Cluster sampling selects subjects from organic groups or clusters of the population for analysis. This

allows researchers to focus their efforts on one or more clusters (like specific schools) and generalize their results to the entire population. For cluster sampling to work, the clusters that researchers study should be similar to the larger population both in terms of matching its characteristics and having internal variance or diversity within the cluster. For instance, if you want to study the home lives of students in a state, you might go door-to-door in a few random neighborhoods that are representative of the state at large. This requires researchers to first identify clusters that might be appropriate for analysis and then to randomly select which cluster(s) to study.

Most applications of cluster sampling are geographic in nature when researchers are doing work on-the-ground and cannot feasibly travel to every school, city, or state in their population. Instead, they will identify schools, cities, or states that are representative of the larger population and then randomly sample from this list to reduce travel time and other difficulties inherent in shifting between contexts.

The major benefit of cluster sampling is feasibility because it allows researchers to study subjects as organic clusters in finite settings. However, for results to be valid, clusters must be representative of the population, and results are further problematized if clusters vary in size and internal diversity or variance from one another.

Census
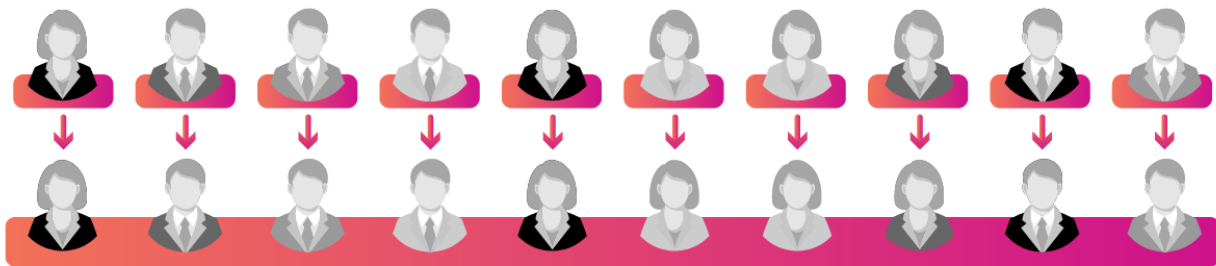
And finally, a census is a study in which all members of a population are treated as subjects. It is not a sampling strategy *per se* but is rather any study where the population (*N*) and the sample (*n*) are identical. Census studies are not very common in education research mainly because it's often difficult to study large groups (e.g., all K-12 teachers in a country), but in situations where populations are relatively small, a census study might be appropriate.

For instance, if you are studying a database of 100 lesson plans and want to be able to determine what percent of the lesson plans in the database have particular characteristics, like learning objectives or standards alignment, then your population would be relatively small (*N* = 100). Even assuming that you aren't trying to generalize to another set of lesson plans, sampling a large enough number to generalize to the database would require you to analyze about 92 lesson plans to achieve a confidence interval of +/-3% at the 95% confidence level (or *n* = 92% of the population). But in a situation like this, if you are already going to analyze 92% of the population, then why not just go ahead and analyze the remaining 8%?

When it is possible to do a census study, doing so provides clear benefits to other sampling procedures, because you can avoid sampling errors and can rely upon basic descriptive statistics (like

mean values) without having to utilize additional statistical analyses for generalization (that could introduce errors).

Considerations

Before choosing a sampling method, you should consider at least two aspects of your study that will determine appropriateness. These include (a) the goal of your research and (b) the contextual feasibility of your study and questions. I'll now explain each of these in more depth.

Research Goal

First, research is done for a variety of purposes. Sometimes we do research to determine norms or trends in a large population, such as whether U.S. legislator mandates for high-stakes testing negatively impact student wellbeing. At other times, we do research to gain a deeper understanding of a phenomenon, such as what it is like to be an undocumented immigrant student in a K-12 school while learning English as a second language. And at other times, we do research to improve a product or intervention, such as iteratively testing a reading intervention program for fourth-grade students. In each of these scenarios, our goal for doing research is different, and our sampling method should reflect this intended goal.

If iterative improvement is our goal, then we will often rely upon sampling methods that are fast and easy or that allow us to answer targeted, finite questions for improvement, thereby favoring convenience or purposeful sampling approaches. If our goal is to understand a phenomenon in-depth (and to remove noise from the larger population), then we will use approaches that are more focused or purposeful. And if our goal is to generalize or to get a broad understanding, then we will use approaches that randomize sampling or that allow us to study every member of the population via a census.

To drive this point home a bit more, many researchers initially believe that randomization is always valuable for research, but randomization is only a tool for attempting to achieve generalization from the sample to the population. It is only appropriate if generalization is the goal. Conversely, if convenience or purposeful sampling is utilized in a study, then generalization can never be achieved. Thus, researchers need to be sure that their method of sampling aligns with their research goals, and the decision of how to sample must be driven by that goal.

Contextual Feasibility

And second, the context of our study and our questions matter in determining sampling methods. Some topics of study might be simpler to observe, measure, or identify, such as a family's income level, while other topics might be more complex and ephemeral, such as a child's emotional wellbeing. If the topic of your study is simpler and easier to observe in a single subject, then you can rely upon approaches that allow for greater numbers of subjects and randomization, but if your topic is more complex, then you will need to use approaches that are more purposeful.

Albert Einstein hinted at this when he famously quipped that "Not everything that counts can be counted, and not everything that can be counted counts." As an example, let's say that we work for a state that has called for schools to submit grant proposals. If we want to study whether the proposals meet specific formatting requirements, we could scan each document with relative ease to see if the

proper headings, font size, and margins were used; or we could write a simple program to do this for us. If, however, we were tasked with evaluating the proposals on scholarly merit or likelihood of success, then this would require much deeper reading of the proposals and would rely upon much more expertise and interpretation on our part. In this scenario, a single researcher could feasibly study formatting issues of a sample of 1,000 proposals with relative ease, but determining each proposal's implementation success likelihood or evidentiary merit would take much more time and effort.

This complexity is also influenced by the field's prior work and knowledge regarding our research topic. If we are studying a topic that is brand new and upon which little previous work exists in terms of theoretical constructs, measurements, tools of observation, and so forth, then we will often need to rely upon more purposeful and focused approaches that allow us to go deep to unpack nuances. If, however, our topic has been studied deeply already and others have provided us with lenses and tools for measuring and observing, then we can use sampling approaches that take a broader view, such as randomization, because we will have a clearer idea of what we are looking for and how to go about finding it.

This issue of feasibility is further complicated in the education research setting because education research is a social science that deals with intentional, agentic humans with rich, complex life histories, which means that the complexity of our study context will also be informed by how we view our subjects and what we expect from them in the course of the study. If we are merely studying external, observable characteristics of our subjects, then larger samples and randomization will be more possible, but if we are expecting to unpack or reveal deep experiences, beliefs, or interpretations, such as trying to understand traumatic experiences among young children, then we will need to be much more purposeful and directed in how we select and interact with our subjects and come to view them not merely as subjects but as participants, informants, or collaborators of the research.

Thus, if we are studying simple topics, we can more feasibly rely upon large sample sizes and randomization to make generalizable determinations, but if we are studying more complex topics, then we may need to reduce sample sizes and focus our efforts on purposefully targeting specific subjects for deeper analysis.
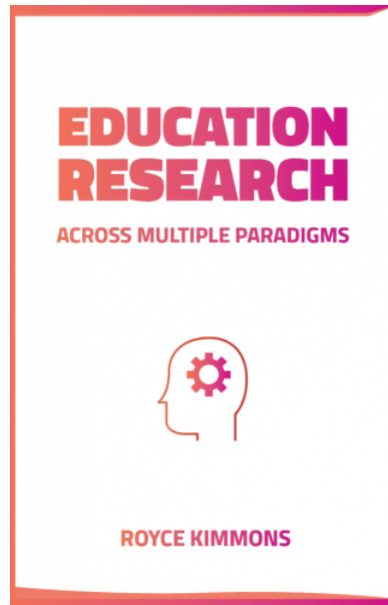
Kimmons, R. (2022). Sampling: How should I choose a sample for my study? In R. Kimmons (Ed.), *Education Research*. EdTech Books. https://edtechbooks.org/education_research/sampling