academicresearchJournals

Vol. 7(8), pp. 254-296, November 2019 DOI: 10.14662/IJELC2019.200

ISSN: 2360-7831
http://www.academicresearchjournals.org/IJELC/Index.htm

International Journal of English Literature and Culture

Full Length Research

Research Paradigms in English Language Teaching

Abera Tsegaye

Wolkite University, Ethiopia. E-mail: tsegayea96@gmail.com

Accepted 4 November 2019

English language and English language teaching (ELT) have been receiving considerable attention from the respective world governments due to their increasing importance for globalization. As a result, ELT research has taken huge strides over the past few decades, which has had its positive implications for improving teaching this very important international language in many parts of the world. However, this has not been the case with and ended up sharing significant common problems pertinent to ELT, which have serious negative implications for these countries' national development, but which also constitute a platform for collaborative research. The establishment of a Research council (RC) in each of the forty five public universities is thus discussed as a possible solution to overcome this problem. The discussion outlines certain challenging and demanding responsibilities for the instructors to help advance ELT research amongst the respective public universities cooperatively and collaboratively specifically where they are teaching in reviewing research materials for the learners to advance the concept of the research.

Key words: Research, English Language Teaching, research paradigms.

Cite This Article As: Abera T (2019). Research Paradigms in English Language Teaching. Inter. J. Eng. Lit. Cult. 7(8): 254-296

Research Paradigms/World Views

The nature of being and Epistemology

These two words always set pupils off philosophy for life. *Ontology* and *epistemology* are the two basic qualities of a branch of philosophy called metaphysics. Metaphysics is anxious with two core questions. 1st, what are the lit of fact? In other way, what are the lit of things that exist? Or what are the universal characteristics of things that exist? These are ontological interrogative. The 2nd qualities of metaphysics are the question, “How can we know the things that exist?” This is an epistemological question. Perhaps, these clarifications are too executive summary and ephemeral to be of much use when the knowledge is new to you. Ontology is anxious with the nature of fact, and different ontological places reflect different directions of what can be real and what cannot. For instance, someone who takes a materialist ontological position (e.g., all that is real is the physical or material world) would ignore the issue that ghosts or spirits can affect the physical world. Why? because ghosts cannot exist whether all that is real is physical. Materialism is one of the core ontological situations, and it is the foundation for much of the research studied in the natural sciences. Nevertheless, a competing view of reality is idealism, which functions that reality is mental and spiritual rather than material (Craig, 1998). Another ontological nature is metaphysical subjectivism. Proponents of that position behave that cognition, what we accept through our senses, creates the state of things as they real exist as opposed to an idealistic than what is in our heads. That is, there is no quality of being lifelike other than what humans produce in their own minds.

You can see how various ontological positions can lead to very different positions on many cases. Recently, for instance, there are arguments in the field of medical informatics that come down to if medical words and diagnoses reflect a materialistic reality that is external to the mind and that exists conscious of the mind versus a more subjective view that they are constructions of the minds of medical specialists (Vicente, Doyle, Milgram, & Burns, 2001). Most of Western natural science is based on modern versions of Descartes's dualism, the idea that both material and mental entities exist. Epistemology is anxious with what we can know about reality (however that is defined) and how we can know it.

At the problem of oversimplification (generalization), ontology refers hat can exist or what is real, and epistemology is about knowledge. In fact, the English word originates from the Greek word episteme, which means "knowledge."

When you raise questions such as "What is knowledge?" "How do I get knowledge?" "How can I be exact of my knowledge (if I can at all)?" and "What are the sets of human knowledge?" you are asking epistemological questions. Epistemology is a critical foundation for research in both the natural and the social sciences. The traditional scientific technique, for instance, is relay on an empirical epistemology: You can understand to know about the world (which, ontologically, is a physical or material world) through properly done experiments. A choice is feminist epistemology, which a debate that much of the research in the social sciences has been studied from a male expects. To a feminist epistemology this is necessary because all knowledge is positioned in the experiences and situation of the researcher. Thus, the knowledge created by a male-exceeds sociology or anthropology will not be equivalent the knowledge created by female-oriented researchers. Hence, feminist epistemologies is focused on a more subjective ontology and also ignore the issue that research is a means of coming to know what is objectively "real." Instead, the knower is always organized by her or his contexts, and thus all knowledge is also positioned (Harding, 1998).

Philosophers have been arguing the idea about involves with the nature of being and involves about the theory of knowledge especially with referred to its methodology, validity and delimitation. Thus, we should think about the views we make about the world when we study research.

Definition of Paradigm

Chalmers (1982) defines a paradigm as "made up of the general theoretical views and laws, and methods for their application that the members of a particular scientific society adopt" (p. 90). Chalmers (1982, p. 91) points out that a paradigm has five components:

- Explicitly stated laws and theoretical assumptions.
- Standard ways of applying the fundamental laws to a variety of situations.
- Instrumentation and instrumental techniques that bring the laws of the paradigm to bear on the real world.
- General metaphysical principles that guide work within the paradigm.
- General methodological prescriptions about how to conduct work within the paradigm.

Skeleton that shows guide and practice of research approach shift of paradigms: Today, in the social sciences, there are several competing paradigms. Some discussions are organized around the view that there are two paradigms, quantitative and qualitative, but that is an oversimplification that shows information rather than foundational beliefs and assumptions. The exact number of world views (paradigms) and the names associated with a particular paradigm vary from author to author, but one generally accepted list includes **three paradigms** (Gephart, 1999; Greene, Benjamin, & Goodyear, 2001; Guba, 1990; Smith, 1989):

- Post-positivism
- Critical theory
- Interpretive

A paradigm is not just a philosophy of science, such as post positivism. It is also the related social science theory, such as behaviorism, and the associated research framework. Finally, it is the application of that entire framework to practice. Each level influences and is influenced by all the other levels. At the basic or fundamental level there is a philosophy of science that makes a number of assumptions about fundamental issues such as the nature of truth (ontology) and what it means to know (epistemology). Although many researchers and practitioners ignore this foundational layer of assumptions, it is an essential aspect of a paradigm.

That is because positivism or post positivism is the philosophy of science that is the foundation for these paradigms. Their methods and practices are based on the assumption that positivism, or at least post positivism, is the true and correct way to look at the world.

And the use of the scientific method as a framework for conducting research would not make sense unless proponents adopt a realist. Ontology: “Reality exists ‘out there’ and is driven by immutable (unchangeable) natural laws and mechanisms. Knowledge of these entities, laws, and mechanisms is conventionally summarized in the form of time (Guba, 1990, p.20). Besides, there are research skeleton such as cooperative study (Heron, 1996; Tan, 2002) that attempt to see problems from multiple perspectives. Such methods would not make much sense without a relativist ontology—“Realities exist in the form of multiple mental constructions, socially and experientially based, local and specific, dependent for their form and content on the persons who hold them” (Guba, 1990, p. 27)—and an assumption that the goal of research is understanding in context instead of the discovery of universal, law like truths.

Positivism

Logical positivism believe that reality is stable and can be observed and described from an objective viewpoint (Levin, 1988), i.e. without interfering with the phenomena being studied. They contend (argue/ believe) that phenomena should be isolated and that observations should be repeatable. This often involves manipulation of reality with variations in only a single independent variable so as to identify regularities in, and to form relationships between, some of the constituent elements of the social world. Predictions can be made on the basis of the previously observed and explained realities and their inter-relationships.

"Positivism has a long and rich historical tradition. It is so embedded (rooted) in our society that knowledge claims not grounded in positivist thought are simply dismissed as a scientific and therefore invalid" (Hirschheim, 1985, p.33). This view is indirectly supported by Alavi and Carlson (1992) who, in a review of research articles, found that all the empirical studies were positivist in approach. Positivism has also had a particularly successful association with the physical and natural sciences. There has, however, been much debate on the issue of whether or not this positivist paradigm is entirely suitable for the social sciences (Hirschheim, 1985). Many authors calling for a more pluralistic attitude research methodologies (see e.g. Kuhn, 1970; Bjørn-Andersen, 1985; Remenyi and Williams, 1996).

In the past four years, French philosopher Auguste Comte established positivism in Western philosophy. He believed societies passed through three stages of explanation. In the first and least enlightened stage, theological explanations dominate. In the second and more enlightened stage, metaphysical or philosophical explanations emerge. For example, a person might say that a particular drug puts people to sleep because it has “durative powers.” Certain characteristics are attributed to items and treated as explanations in the metaphysical stage. And in the third and highest stage, positivism, scientific explanations are the rule. Comte advocated the emerging sciences, such as astronomy, biology, physics, and chemistry, but he was also a founder of sociology and was concerned that this field of human study should be based on a solid scientific foundation. He sometimes called sociology “social physics” and argued that the methods that were so successful in the natural sciences should also be applied to the human sciences. He advocated the use of the scientific method to validate theories of human behavior: Scientifically speaking, all isolated, empirical observation is idle, and even radically uncertain; science can use only those observations which are connected, at least hypothetically, with some law; that it is such a connection which makes the chief difference between scientific and popular observation, embracing the same facts, but contemplating(considering) them from different points of view: and that observations empirically conducted can at most supply provisional materials, which must usually undergo an ulterior revision. The observer would not know what he ought to look at in the facts before his eyes, but for the guidance of a preparatory theory. This is undisputed with regard to astronomical, physical, and chemical research, and in every branch of biological study. Carrying on the analogy, it is evident that in the corresponding divisions of social science, there is more need than anywhere else of theories which shall scientifically connect the facts that are happening with those that have happened. (Comte, 1854)

In the real practice/finding/result from the study, science was seen as the way to get at truth, to understand the world well enough so that we might predict and control it. The world and the universe were deterministic -- they operated by laws of cause and effect that we could discern (distinguish) if we applied the unique approach of the scientific method. Science was largely a mechanistic or mechanical affair. We use deductive reasoning to postulate theories that we can test. Based on the results of our studies, we may learn that our theory doesn't fit the facts well and so we need to revise our theory to better predict reality.

The positivist believed in *empiricism* -- the idea that observation and measurement was the core of the scientific endeavor. The key approach of the scientific method is the experiment, the attempt to discern natural laws through direct manipulation and observation.

Critical Theory

One of the most common forms of post-positivism is a philosophy called *critical realism*. A critical realist believes that there is a reality independent of our thinking about it that science can study. (This is in contrast with a *subjectivist* who would hold that there is no external reality -- we're each making this all up!). Positivists were also realists. The difference is that the post-positivist critical realist recognizes that all observation is fallible (imperfect) and has error and that all theory is revisable. In other words, the critical realist is *critical* of our ability to know reality with certainty. Where the positivist believed that the goal of science was to uncover the truth, the post-positivist critical realist believes that *the goal of science is to hold steadfastly (persistently) to the goal of getting it right about reality, even though we can never achieve that goal*!

Because all measurement is fallible, the post-positivist emphasizes the importance of multiple measures and observations, each of which may possess different types of error, and the need to use *triangulation* across these multiple errorful sources to try to get a better bead on what's happening in reality. The post-positivist also believes that all observations are theory-laden besides which scientists (and everyone else, for that matter) are inherently biased by their cultural experiences, world views, and so on. This is not cause to stop in despair; however, Just because I have my world view based on my experiences and you have yours doesn't mean that we can't hope to translate from each other's experiences or understand each other. That is, post-positivism rejects the *relativist* idea of the *incommensurability* of different perspectives, the idea that we can never understand each other because we come from different experiences and cultures. Most post-positivists are *constructivists* who believe that we each construct our view of the world based on our perceptions of it. Because perception and observation is fallible, our constructions must be imperfect. So what is meant by *objectivity* in a post-positivist world? Positivists believed that objectivity was a characteristic that resided in the individual scientist.

Scientists are responsible for putting aside their biases and beliefs and seeing the world as it 'really' is. Post-positivists reject the idea that any individual can see the world perfectly as it really is. We are all biased and all of our observations are affected (theory-laden). Our best hope for achieving objectivity is to triangulate across multiple fallible perspectives! Thus, objectivity is not the characteristic of an individual; it is inherently a social phenomenon. It is what multiple individuals are trying to achieve when they criticize each other's work. We never achieve objectivity perfectly, but we can approach it. The best way for us to improve the objectivity of what we do is to do it within the context of a broader contentious (debatable) community of truth-seekers (including other scientists) who criticize each other's work. The theories that survive such intense scrutiny are a bit like the species that survive in the evolutionary struggle. (This is sometimes called the *natural selection theory of knowledge* and holds that ideas have 'survival value' and that knowledge evolves through a process of variation, selection and retention). They have adaptive value and are probably as close as our species can come to being objective and understanding reality.

Interpretivism

Interpretivists intended that only through the subjective interpretation of and intervention in reality can that reality be fully understood. The research work of act in their natural compound is key to the interpretive philosophy, along with the acknowledgement that scientists cannot destroy affecting those phenomena they acquiring knowledge. They admit that there may be many interpretations of reality, but maintain that these interpretations are in themselves a part of the scientific knowledge they are pursuing. Interpretive does not have a tradition that is no less glorious than that of positivism, nor is it shorter.

This approach to social science research rejects the positivist idea that the same research methods can be used to study human behavior as are successfully used in fields such as chemistry and physics. Interpretivists argue that when you study the behavior of a metal, the primary causes of changes in the metal are in the environment (e.g., heat, stress). Humans behave the way they do in part because of their environment. However, that influence is not direct as it is with a piece of metal.

Humans are also influenced by their subjective perception of their environment—their subjective realities. We do not worry about the subjective impressions of a steel bar, but if we are to fully understand the behavior of an 18-year-old delinquent we must understand her view of the world around her. We must also understand the subjective perceptions of her by others in her social and cultural context. Thus, for interpretivists, what the world means to the person or group being studied is critically important to good research in the social sciences. Interpretivists favor qualitative methods such as case studies, interviews, and observation because those methods are better ways of getting at how humans interpret the world around them. Some of the philosophical foundations of interpretive can be found in Immanuel Kant's Critique of Pure Reason (1781/2003), in which he argued that humans interpret their sensations; they do not directly experience the "out there" world as it is.

Constructivism

The constructivist paradigms elongate interpretive philosophy by showing the necessity of explaining how unique participants in a social setting construct their beliefs (Guba and Lincoln 1989). It supply specific due to the different achievement of the one who conduct a research and other participants in a research time and place and then needs to develop a consensus within several members about how to perceive the focus of study. The constructivist research report will highlight unique ideas of the community program or other ideas and describe how a consensus can be reached within members.

Constructivist study benefit an reactive research progress, in which someone who conduct a research starts an evaluation in some group setting by differentiating the variety needs society in that time and place.

The researcher goes on to learn what each group thinks and then slowly tries to develop a shared perspective on the problem being evaluated (Guba and Lincoln 1989).

The constructions of a variety of individuals—deliberately chosen so as to uncover widely variable viewpoints—are elicited, challenged, and exposed to new information and new, more sophisticated ways of interpretation, until some level of consensus is reached (although there may be more than one focus for consensus). (Guba and Lincoln 1989:180–81)

The researcher conducts an open-ended interview with the first respondent (R1) to learn about her thoughts and attitudes on the subject of the study, her "construction" (C1). The researcher then asks this respondent to select a second respondent (R2), who perceive very slightly. The second respondent is then interviewed in the similar but also asked to comment on the titles raised by the previous respondent. The process continues until all major issues are represented, and it may be repeated with the same set of respondents.

Postmodernism is an umbrella term that is used in different ways by different speakers. As a result, this term defies easy summary. Typically, however, speakers invoking "postmodernism" are committed to exploring the complex relationships of power, knowledge, and discourse created in the struggle between social groups. Additionally, postmodernism is intertwined with several other perspectives that challenge the conduct of business as usual. These traditions include feminism (Mumby, 1996), neo- Marxism, post structuralism (Parker, 1995), and post colonialism).

The terms modernity and postmodernist as descriptors of the contemporary period are characterized by dramatic changes in global politics, economics, and culture. These changes have also been described as postmodernism.

Needless to say, the experience of modernity for its affected groups has been signed by fluctuation and contradiction.

These experiences include transformation (e.g., following the removal of traditions and the sensing of new chances for identity), organization (e.g., recruitment, integration, and compulsory performance within institutions), differentiation (e.g., the simultaneous management of multiple commitments to various groups), and reflection (e.g., nostalgia and hope created by differences between the status quo and actual or imagined alternatives).

Postmodernism, in other ways, describes a series of breaks and continuities between modern and contemporary conditions. Although the relationship between modernity and post modernity is often predict as a dichotomy, this picture is not useful. It shows that each entity is independent and monolithic, when in fact both are marked by contingency and variety. For example, observers have categorized varieties of postmodernism that differ based on their proponents' reaction to changing conditions. Some postmodernists, for example, affirm and embrace change (although critics have noted that this stance has been appropriated for questionable ends, as in the "creative destruction" of organizational re-engineering). Other postmodernists are more skeptical, seeking to direct change toward the subversion of modern rationality. In this way, it's best to think of modernism and postmodernism as existing in a mutually constitutive relationship (Mumby, 1997). Neither form of life is separate or total; each contains the seeds and residues of the other. In fact, each requires the continued existence of the other in order to appear—through opposition—distinct and logically sequenced.

In their recent survey of postmodernism, Best and Kellner (2001) argue that “the transition to a postmodern society is bound up with fundamental changes that are transforming pivotal phenomena from warfare to education to politics, while reshaping the modes of work, communication, entertainment, everyday life, social relations, identities, and even bodily existence and life-forms.’’

This is obviously a very broad field to survey. Some of its most famous explorers (many of whom are European) include Baudrillard (1994), Bhaba (1994), Deleuze and Guattari (1987), Derrida (1976, 1978), Foucault (1972, 1973a, 1973b, 1979), Gergen (1991), Jameson (1983), Laclau and Mouffe (1985), Lyotard (1984), Rorty (1989), and Said (1983).

These primary commentators—and their numerous followers and interpreters— have consistently noted particular conditions that bridge the 20^{th} and 21st centuries (Foster, 1983). If you have found this list disorienting or disturbing, you’re not alone. Postmodernism is, maddeningly, both urgent and playful. It uses the strategies of blankness, irony, and reflexivity to heighten our awareness of paradox, ambiguity, uncertainty, emergence, and difference.

It works similar to a various to disrupt the tendencies of modernist thinking that “turn verbs into nouns, process into structure, relationships into things and constructs into concrete (reified) objects” (Chia, 1995, p. 589). It reminds us that our awareness and identities—all of the taken-for-granted elements of human organization—might have been, and might yet be; otherwise, These elements, which we may have viewed as total, transcendent, or fixed, are unknowingly vulnerable (Mumby, 1997). They are “problematized”—recovered for the function of interrogation (cross-examination), critique, and transformation.

RESEARCH AND THE RESEARCH PROBLEM

Purposes/intention

• To describe the nature, criteria of/for, and the different types of research approach.
• To express some aspects/actions of the debate about the nature of knowledge and the value of scientific method.
• To announce the idea at the heart of any research project, the research problem, and to discuss research problem.
• To give common mistakes.
• To explain how a research problem is found and stated.

THE RESEARCH APPROACH

The Oxford Encyclopedic English Dictionary defines research in many ways; some of these are:

- The systematic investigation into the study of materials, sources etc. in order to establish facts and reach new conclusions.
- An endeavor to discover new or collate old facts etc. by the scientific study of a subject or by a course of critical investigation.

Leady defines it from a more utilitarian point of view:

- Research is a procedure by which we attempt to find systematically, and with the support of demonstrable fact, the answer to a question or the resolution of a problem. (1989, p. 5)

Domino ski is so terse in his definition that he seems to miss the point --research is a fact-finding activity. (1980, p. 2)

- Kerlinger uses more technical language to define it as: the systematic, controlled, empirical and critical investigation of hypothetical propositions about presumed relations among natural phenomena. (1970, p. 8)

- **Research has to be scientific to be of any value:**

• Research has to be objective.
• Research has to involve some quantification.
• Research has to be general sable: a description of a single case has no practical value.

SCIENTIFIC METHOD IN RESEARCH

Scientific method is the discipline which forms the foundation of modern scientific enquiry.
It is therefore important to mention some of the main assumptions made in this method of enquiry, and to describe some of its major characteristics.

Scientific method has been applied, to a greater or lesser extent, to research in some areas not principally thought of as ‘scientific’, such as sociology, psychology and education, although some scientists question the appropriateness of doing this.

According to Cohen and Manion (1994, pp. 12–16) there are five major assumptions:

The first major assumption is the belief that there is some kind of order in the universe, and that it is possible for us to gain some understanding of this order. This is linked with the idea of determinism, the assumption that events have causes, and that the links between events and causes can be revealed. This regularity enables some predictions to be made about future events (e.g. if gravity causes apples to fall today, it will also cause them to fall tomorrow).

The second view is that, in order to enable us to gain this understanding of the world, there must be an agreement between people that outside truth exists, and that people recognize the same reality, a public or shared reality. It is rarely essential to point out that much philosophic debate has been devoted to the nature of reality. Nevertheless, scientific enquiry relies on the acceptance of the reliability of knowledge gained by experience to provide empirical evidence (evidence which is verifiable by observation) to support or refute its theories.

The third assumption is the reliability of human perception and intellect. Despite the many ways in which our senses can be tricked, researchers depend on their senses to record and measure their work reliably. Reasoning is an important method of organizing data and ideas, and is regarded, if used correctly, as a dependable tool of research. Human memory also plays a major role in research. To avoid questioning at every single stage, some credence must be given to the power of memory to provide reliable knowledge.

The fourth assumption is the principle of parsimony. Phenomena should be explained in as economic manner as possible. Needless complexity is abhorred, and scientists aim to achieve the most elegant and simple theories.

The fifth assumption is that of generality. This is the assumption that there can be valid relationships between the particular cases investigated by the researcher and the general situation in the world at large. It is accepted that these relationships can be relatively unproblematic in some sciences (e.g. chemistry and physics) but that in others, with a larger number of unknown factors (e.g. sociology), there is a weaker chance of generality.

Characteristics of research which uses scientific method:

- It necessitates clarification of a goal-- It entails a specific programme of work
- It is intentioned at enhancing cognition by interpreting facts or ideas and reaching some conclusions about their meaning
- It needs reasoned argument to help conclusions
- It is reiterative in its practices.

THE RESEARCH PROBLEM

One of the first tasks, therefore, on the way to deciding on the detailed topic of research is to find a question, an unresolved controversy, a gap in knowledge or an unrequited need within the chosen subject. Here is a list of the most important:

- It should be of great willing to you
- The problem should be important
- It should be delineated
- You should be able to receive the data required
- You should be able to draw conclusions related to the problem
- You should be able to state the problem clearly and concisely.

The problem can be generated either by an initiating idea, or by a perceived problem area. For example, investigation of 'rhythmic patterns in settlement planning' is the product of an idea that there are such things as rhythmic patterns in settlement plans, even if no-one had detected them before. This kind of idea will then need to be formulated more precisely in order to develop it into a researchable problem.

HELPS TO LOCATING AND ANALYSING PROBLEMS

Booth et al. (1995, p. 36) supplies that the means for focusing on the formulation of your research problem looks like this:

1. Find an interest in a broad subject area (problem area).
2. Narrow the interest to a plausible topic.
3. Question the topic from several points of view.
4 . Define a rationale for your project.

We should keep in mind three questions when engaged in the preliminary exploratory work. The first is what is your motivation for doing the research? The second question is, what relevant interest, experience or expertise do you bring to bear on the subject? The third question is, what are you going to produce?

Literature review and defining the problem area

The objective of the initial review of the literature is to discover relevant material published in the chosen field of study and to search for a suitable problem area. Fox (1969) mentions two kinds of literature which should be reviewed. The first is 'conceptual literature'. This is written by authorities on the subject you have in mind, giving opinions, ideas, theories or experiences, and published in the form of books, articles and papers. The second is 'research literature' which gives accounts and results of research which has been undertaken in the subject, often presented in the form of papers and reports.

Definition of research problem

From the interest in the wider issues of the chosen subject, and after the selection of a problem area, the next step is to define the research problem more closely so that it becomes a specific research problem, with all the characteristics already discussed. This stage requires an enquiring mind, an eye for inconsistencies and inadequacies in current theory and a measure of imagination.

Review of related literature considered as secondary information

A more focused review of literature follows the formulation of the research problem. The purpose of this review is to learn about research already carried out into one or more of the aspects of the research problem, in order to: 1 summarize the results of previous research to form a foundation on which to build your own research 2 collect ideas on how to gather data 3 investigate methods of data analysis 4 study instrumentation which has been used 5 assess the success of the various research designs of the studies already undertaken.

TYPES OF RESEARCH

Intention

• To explain the nature of awareness and the way they can be measured.
• To find the mandate of theory.
• To examine unique types of research and their characteristics

Lots of research questions which instigate research contain approaches to research which are distinguished by their theoretical background and methodologies. A brief summary of various types of research will illustrate the possibilities for your research efforts. Several major types of research can be identified.

Back ground of the research

Historical research has been defined as the systematic and objective location, evaluation and synthesis of evidence in order to establish facts and draw conclusions about past events (Borg, 1963).

It involves exploring the meaning and relationship of events, and as its resource it uses primary historical data in the form of historic artifacts, records and writings. It attempts to find out what happened in the past and to reveal reasons for why and how things happened. The value of historical research has been categorized by Hill and Kerber (1967) as:

- It enables solutions to contemporary problems to be sought (required) in the past.
- It throws light on present and future trends.
- It stresses the relative importance and the effects of the interactions that are found within all cultures.
- It allows for the revaluation of data supporting selected hypotheses, theories and generalizations that are presently held about the past

Historical evidence, consisting of primary historical data, must be scrutinized from two points of view. The first is to ascertain whether the artifact or document to be studied is genuine. The second is to examine, in written evidence in the form of historic documents etc. What is the meaning of what is written, and how accurate is it? For example, many authentic medieval texts are known to be wildly inaccurate and vague in their descriptions of events.

❖ As Gottschalk (1951) there are four aspects of historical research which determine the scope of a study:

❖

✓ where the events take place
✓ which people are involved
✓ when the events occurred
✓ What kind of human activity was involved?
✓ The degree to which an aspect is studied can be varied, i.e. the number of human activities examined can be increased or decreased, the time-span covered can be extended or contracted etc. It must be remembered that the mere collection of historic facts, or the setting up of chronologies of events, does not constitute research. Although these are a necessary part of historical research, an interpretation of the meanings and an assessment of the significance of the events are required.

Historical research is not based purely on scientific method. For instance, the data used are seldom based on direct observation or experimentation. But it should share many of the disciplines of scientific method, such as objectivity and the desire to minimize bias and distortion, the use of scientific techniques such as chemical and radioactive analysis, and statistics. The problem for historians tends to be the paucity (scarcity) of information, while scientists are often overwhelmed by it!

COMPARATIVE

Comparative research is often used together with historical research. Researchers compare people's experience of different societies, either between times in the past or in parallel situations in the present. These studies can be on the macro level, e.g. studying the role of revolutions in class struggle, or on the micro level, e.g. individual experiences in different types of marriage. It is often easier to understand phenomena when they are compared with similar phenomena from another time or place. Culture and society rely heavily on what has gone before and often use references from the past to justify the present. The constitutions, the tax system, social mores (way of living) are all rooted in their own histories. Similarly, place also determines that phenomena develop differently. The study and comparison of differences help to reveal the origins and development of social phenomena, locating them in a certain time and place, and thus defeating claims that they are universal and a temporal.

Many social theories are presented as if the generalizations that they embody are valid for all times and places, when in fact they were arrived at on the basis of limited contemporary Western experience. (Llobera, 1998, p. 74) We can also learn by making comparisons both with the past and with experiences elsewhere. It would be foolish for politicians to introduce, say, sweeping changes to the electoral system, without carefully studying the effects of such changes in the past and in other situations.

Explanatory

Instead of examining record or artifacts, descriptive research relies on observation as a means of collecting data. It attempts to examine situations in order to establish what is the norm, i.e. what can be predicted to happen again under the same circumstances. ‘Observation’ can take many forms.
The scale of the research is influenced by two major factors:

1. The level of complexity of the survey.
2. The scope of the survey.

For example, seeking relationships between specific events inevitably requires a more complex survey technique than aiming merely to describe the nature of existing conditions. Likewise, surveying a large number of cases over a wide area will require greater resources than a small, local survey.

In order both to save on unnecessary work and to give accurate information on the subject of the research, the sample of people or events surveyed (technically called the population) must be carefully chosen and delineated (described/explained). To do this, it is necessary to be aware of the precise subject focus of the research so that specific objectives can be formulated. As descriptive research depends on human observations and responses, there is a danger that distortion of the data can occur. This can be caused, among other ways, by inadvertently (unintentionally) including biased questions in questionnaires or interviews, or through selective observation of events. Although bias cannot be wholly eliminated, an awareness of its existence and likely extent is essential.

CORRELATION

The information sought in correlation research is expressed not in the form of artifacts, words or observations, but in numbers. While historical and explanatory approaches are predominantly forms of qualitative research, analytical survey or correlation research is principally quantitative. ‘Correlation’ is another word to describe the measure of association or the relationships between two phenomena. In order to find meaning in the numerical data, the techniques of statistics are used. What kind of statistical tests are used to analyze the data depends very much on the nature of the data. This form of quantitative research can be broadly classified into two types of studies: relational studies, and prediction studies.

Relational studies is an investigation of possible relationships between phenomena to establish if a correlation exists and, if so, its extent. This exploratory form of research is carried out particularly where little or no previous work has been done, and its outcomes can form the basis for further investigations.

Prediction studies tend to be carried out in research areas where correlations are already known. This knowledge is used to predict possible future behavior or events, on the basis that if there has been a strong relationship between two or more characteristics or events in the past, then these should exist in similar circumstances in the future, leading to predictable outcomes.

In order to produce statistically significant results, quantitative research demands data from a large number of cases. Greater numbers of cases tend to produce more reliable results; 20–30 is considered to be about the minimum, though this depends on the type of statistical test applied. The data, whatever their original character, must be converted into numbers.

One of the advantages of correlation research is that it allows for the measurement of a number of characteristics (technically called variables**)** and their relationships simultaneously. Particularly in social science, many variables contribute to a particular outcome (e.g. satisfaction with housing depends on many factors). Another advantage is that, unlike other research approaches, it produces a measure of the amount of relationship between the variables being studied. It also, when used in prediction studies, gives an estimation of the probable accuracy of the predictions made.

One limitation to what can be learned from correlation research is that, while the association of variables can be established, the cause and effect relationships are not revealed.

EXPERIMENTAL

Experimental research, where the researcher can artificially control causal factors, is not really possible in social research. However, the idea is put forward that history and comparison can often supply the researcher with what is a natural experiment. According to Mill's method of agreement (one of his five 'methods of experimental enquiry' devised in the nineteenth century), 'If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstances in which alone all the instances agree is the cause (or effect) of the given phenomenon' (Mill, 1973, p. 390).

These types of research differ from the other research approaches noted above through its greater control over the objects of its study. The researcher strives to isolate and control every relevant condition which determines the events investigated, so as to observe the effects when the conditions are manipulated. Chemical experiments in a laboratory represent one of the purest forms of this research type.

In other way round, an experiment involves making a change in the value of one variable – called the independent – and observing the effect of that change on another variable – called the dependent variable. (Cohen and Manion, 1994)

Thus, the most important characteristic of the experimental approach is that it deals with the phenomenon of 'cause and effect'. However, the actual experiment is only a part of the research process. There are several planned stages in experimental research.

This allows decisions to be made as to what variables are to be tested and how they are to be controlled and measured. This stage, called the design of the experiment, must also include the choice of relevant types of test and methods of analyzing the results of the experiments (usually by statistical analysis). Pre-tests are then usually carried out to detect any problems in the experimental procedure.

Only after this is the experiment proper carried out. The procedures decided upon must be rigorously adhered to and the observations meticulously recorded and checked. Following the successful completion of the experiment, the important task – the whole point of the research exercise – is to process and analyze the data and to formulate an interpretation of the experimental findings. Not all experimental research has to, or even can, take place in a laboratory. The experimental methods used must take account of how much it is possible to control the variables. Writers of textbooks on research have classified experimental designs in different ways. Campbell and Stanley (1966) make their categorization into four classes: 1 pre-experimental 2 true experimental 3 quasi-experimental 4 correlations and ex post facto.

Things done before experimental designs are unreliable and primitive experimental methods in which assumptions are made despite the lack of essential control of variables. An example of this is the supposition that, faced with the same stimulus, all samples will behave identically to the one tested, despite possible differences between the samples.

True experimental designs are those which rigorously check the identical nature of the groups before testing the influence of a variable on a sample of them in controlled circumstances.

Parallel tests are made on identical samples (control samples) which are not subjected to the variable. In quasi-experimental designs, not all of the conditions of true experimental design can be fulfilled. The nature of the shortcomings is however recognized, and steps are taken to minimize them or predict a level of reliability of the results.

The most common case is when a group is tested for the influence of a variable and compared with a non-identical group with known differences (control group) which has not been subjected to the variable. Another, in the absence of a control group, is repeated testing over time of one group, with and without the variable (i.e. the same group acts as its own control at different times).

Correlation design looks for cause and effect relationships between two sets of data, while ex post facto designs turn experimentation into reverse, and attempt to interpret the nature of the cause of a phenomenon by the observed effects. Both of these forms of research result in conclusions which are difficult to prove and they rely heavily on logic and inference.

EVALUATION

This is a descriptive type of research specifically designed to deal with complex social issues. It aims to move beyond 'just getting the facts' in order to make sense of the myriad human, political, social, cultural and contextual elements involved. The latest form of this type of research, named by Guba and Lincoln as fourth-generation evaluation, has, according to them, six properties:

1. The evaluation outcomes are not intended to represent 'the way things really are, or how they work', but present the meaningful constructions which the individual actors or groups of actors create in order to make sense of the situations in which they find themselves.
2 .In representing these constructions, it is recognized that they are shaped to a large extent by the values held by the constructors. This is a very important consideration in a value-pluralistic society, where groups rarely share a common value system.
3 .These constructions are seen to be inextricably linked to the particular physical, psychological, social and cultural contexts within which they are formed and to which they refer. These surrounding conditions, however, are themselves dependent on the constructions of the actors which endow them with parameters, features and limits.
4. It is recognized that the evaluation of these constructions is highly dependent on the involvement and viewpoint of the evaluators in the situation studied.
5. This kind of research exerted that evaluation should be action oriented, define a course which can be practically followed, and stimulate the carrying out of its recommendations. This usually requires a stage of negotiation with all the interested parties.
6 .Due regard should be given to the dignity, integrity and privacy of those involved at any level, and those who are drawn into the evaluation should be welcomed as equal partners in every aspect of design, implementation, interpretation and resulting action. (1989, pp. 8–11)

There are a range of different approaches or evaluation models. Two of them are systems analysis and responsive evaluation.

Systems analysis is a holistic type of research, which reverses the three stage order of thinking which is typical of scientific enquiry, i.e. breaking the problem or phenomenon to be investigated down into researchable parts, then separately evaluating the parts, and finally aggregating these evaluations into an explanation of the whole. In systems analysis, there are also three stages, but they start from appraising the whole: 1 identifying an encompassing whole (system) of which the phenomenon or problem is a part 2 evaluating the behavior or properties of the encompassing whole 3 explaining the behavior or properties of the phenomenon or problem in terms of its roles or functions within the encompassing whole. Systems analysis lends itself to creating understanding in complicated situations, particularly those involving people and organizations; such problems are often referred to as 'messes' because of their indeterminate nature and large number of interconnected variables. Modeling and diagramming are two of the principal techniques used to describe systems.

In the responsive evaluation model a series of investigative steps is undertaken in order to evaluate how responsive a programme is (e.g. an advertising campaign, a new degree course or an experimental traffic scheme) to all those taking part in it, including:

• Data collection Identifying issues from the people directly involved in the programme; identifying further issues from the programme documents; observing how the programme is actually working.
• Evaluation The design of an evaluation based on the data collected and reporting findings.
• Suggesting changes, informing the participants of the findings in ways specifically designed for each type of audience.

A common purpose of evaluation research is to examine programmes or the working of projects from the point of view of levels of awareness, costs and benefits, cost-effectiveness, attainment of objectives and quality assurance. The results are generally used to prescribe changes to improve and develop the situation, but in some cases might be limited to descriptions giving a better understanding of the programme (Robson, 1993, pp. 170–9).

These can be seen as related to experimental research, though it is carried out in the real world rather than in the context of a closed experimental system. A basic definition of this type of research is: 'a small scale intervention in the functioning of the real world and a close examination of the effects of such an intervention' (Cohen and Manion, 1994, p. 186).

Its main characteristic is that it is essentially an 'on the s spot' procedure, principally designed to deal with a specific problem evident in a particular situation. No attempt is made to separate a particular feature of the problem from its context in order to study it in isolation. Constant monitoring and evaluation are carried out, and the conclusions from the findings are applied immediately, and further monitored. Action research depends mainly on observation and behavioral data. As a practical form of research, aimed at a specific problem and situation and with little or no control over independent variables, it cannot fulfill the scientific requirement for generalizability. In this sense, despite its exploratory nature, it is the antithesis of experimental research.

ETHNOGENIC

In this approach, the researcher is interested in how the subjects of the research theorize about their own behavior rather than imposing a theory from outside. The test of success is that the subjects themselves recognize the description of familiar features of their culture. As a process of studying human behavior, according to Goetz and LeCompte (1984), the ethnogeny approach has three characteristic features:

It aims to represent a view of the world as it is structured by the participants under observation by eliciting phenomenological data; it takes place in the undisturbed natural settings of the subjects; and it attempts to represent the totality of the social, cultural and economic situation, regarding the context to be equally important as the action (Uzzell, 1995, pp. 304–5).

This is a difficult form of research for several reasons. As so much of culture is hidden and rarely made explicit, the data being sought by the researcher need to be pursued by delving deep into the language and behavior of the subjects of the study, and of the surrounding conditions in which they live.

There is an ever-present danger that the cultural background and assumptions of the researcher will unduly influence the interpretations and descriptions made on the basis of the data collected. In addition to this, there can be confusions produced by the use of language and the different meanings which may be given to words by the respondents and researcher.

The accounts of events in the past can never capture the infinite contents of history. Historical knowledge, however well authenticated, is always subject to the biases and memory of its chronicler.

Different from these problems of interpretation of data, there is the fact that when working in a naturalistic setting, with social groups engaged in everyday activities, it is impossible to repeat the situation in order to verify the research. Social reality is not stable: a thing never 'is', as it is always changing into something else. It is therefore of great importance that multi-method and confirmatory data sources are used to capture the moment.

FEMINIST

Feminist research is a particular model of social research which involves theory and analysis that highlight the differences between men's and women's lives. It claims that researchers who ignore these differences have invalid knowledge, as non-feminist paradigms usually ignore the partiality of researchers' ideas about the social world. Value neutrality is impossible as no researcher practices research outside his or her system of values and no methods of social science can guarantee that knowledge is originated independently of values.

No specific methods are seen to be particularly feminist, but the methodology used is informed by theories of gender relations. However, feminist research is undertaken with a political commitment to the identification and transformation of gender relations. This tends to reveal that this form of research is not uniquely political, but rather exposes all methods of social research to be political.

CULTURAL

Many of the prevailing theoretical debates (e.g. postmodernism, post structuralism) are concerned with the subjects of language and cultural interpretation, with the result that these issues have frequently become central to sociological studies. The need has therefore arisen for methodologies that allow analysis of cultural texts to be compared, replicated, disproved and generalized. From the late 1950s, language has been analyzed from several basic viewpoints: the structural properties of language (notably Chomsky, Sacks, and Schegloff), language as an action in its contextual environment (notably Wittgenstein, Austin and Searle) and sociolinguistics and the ‘ethnography of speaking’ (Hymes, Bernstein, Labov and many others).

However, the meaning of the term ‘cultural texts’ has been broadened from that of purely literary works to that of the many manifestations of cultural exchange, be they formal such as opera, TV news programmes, cocktail parties etc., or informal such as how people dress or converse. The main criteria for cultural texts are that one should be able to ‘read’ some meanings into the phenomena. Texts can therefore include tactile, visual and aural aspects, even smells and tastes. Three approaches to the consistent interpretation of cultural texts can be mentioned here briefly: content analysis, semiotics and discourse analysis.

Topic analysis was developed from the mid1900s chiefly in America, and is a rather positivistic attempt to apply order to the subjective domain of cultural meaning. A quantitative approach is taken by counting the frequency of phenomena within a case in order to gauge its importance in comparison with other cases. As a simple example, in a study of racial equality, one could compare the frequency of the appearance of black people in television advertisements in various European countries. Much importance is given to careful sampling and rigorous categorization and coding in order to achieve a level of objectivity, reliability and generalizability and the development of theories.

Semiotics takes an almost opposite approach by attempting to gain a deep understanding of meanings by the interpretation of single elements of text rather than to generalize through a quantitative assessment of components. The approach is derived from the linguistic studies of Saussure, in which he saw meanings being derived from their place in a system of signs. Words are only meaningful in their relationship with other words, e.g. we only know the meaning of ‘horse’ if we can compare it with different animals with different features. This approach was further developed by Barthes and others to extend the analysis of linguistic-based signs to more general sign systems in any sets of objects: semiotics as a method focuses our attention on to the task of tracing the meanings of things back through the systems and codes through which they have meaning and make meaning. (Slater, 1995, p. 240)

Hence the meanings of a red traffic light can be seen as embedded in the system of traffic laws, color psychology, codes of conduct and convention etc. (which could explain why in China a red traffic light means ‘go’). A strong distinction is therefore made between denotation (what we perceive) and connotation (what we read into) when analyzing a sign.

Discourse analysis studies the way that people communicate with each other through language within a social setting. Language is not seen as a neutral medium for transmitting information; it is bedded in our social situation and helps to create and recreate it. Language shapes our perception of the world, our attitudes and identities. While a study of communication can be simply broken down into four elements (sender, message code, receiver and channel), or alternatively into a set of signs with both syntactical (i.e. orderly or systematic) organization and semantic (i.e. meaningful and significant) relationships, such simplistic analysis does not reflect the power of discourse.

It is the triangular relationship between discourse, cognition and society which provides the focus for this form of analysis (van Dijk, 1994, p. 122). Two central themes can be identified: the interpretive context in which the discourse is set, and the rhetorical organization of the discourse. The former concentrates on analyzing the social context, for example, the power relations between the speakers (perhaps due to age or seniority) or the type of occasion where the discourse takes place (a private meeting or at a party). The latter investigates the style and scheme of the argument in the discourse, for example a sermon will aim to convince the listener in a very different way to a lawyer’s presentation in court.

Poststructuralist social theory, and particularly the work of the French theorist Michel Foucault, has been influential in the development of this analytical approach to language. According to Foucault, discourses are ‘practices that systematically form the objects of which they speak’ (1972, p. 43). He could thus demonstrate how discourse is used to make social regulation and control appear natural.

Grounded theory

Glaser and Strauss (1967) developed grounded theory as a reaction to the then recent stress on the need to base social research on pre-defined theory. Grounded theory takes the opposite approach – it does the research in order to evolve the theory. This gives rise to a specific style of procedure and use of research methods.

CHAPTER II

RESEARCH METHODS

Intention/purposes

• To show the range of research methods available to the researcher.
• To describe the nature of different research methods used for collecting qualitative and quantitative data.
• To indicate appropriate methods of analyzing data

INTRODUCTION

The intention of this chapter is to provide a very brief description of a range of commonly used research methods so that you have a basic idea of the options open to you when you set about planning your research. An essential ingredient of your research proposal will be to suggest appropriate methods by which you will carry out your research, so a prior understanding of the range of methods open to you will help you to decide which might be pertinent to your form of study. Although the range of methods available for research into the natural sciences, the social sciences and the humanities is enormous, the number of methods which could reasonably be explained in this chapter is limited, so only a general indication can be given of some of the principal methods.

It would, at first glance, be easy to cover the subject by neatly dividing it under two headings: methods of data collection, and methods of data analysis. The first part would be a straightforward summary of how to collect data from different sources using different methods, and the second would be a description of techniques of analysis, both quantitative and qualitative. However, it soon becomes evident that both of the activities of data collection and data analysis tend to be inextricably bound up with the research strategies – historical, survey, case study etc. – and cannot easily be discussed without some reference to these.

Therefore, as an introduction, we will first look briefly at the three perspectives from which one can review research methods, examine the characteristics which make them distinctive, and describe how they interact. The three perspectives are: data collection and analysis, quantitative and qualitative research, and research strategies.

DATA COLLECTION AND ANALYSIS

Once the research problem has been formulated, it should become evident what kind of data will be required to study the problem, and also what kind of analysis will be appropriate to analyze the data. In nut shell, it will often be appropriate to decide first on the type of analysis, quantitative or qualitative, which will be required to investigate your research problem, and then on the type of data which need to be collected in order to make that analysis.

The types of data collection and analysis are not always as distinct as they might at first appear. for example in exploratory research, the data may be continuously analyzed as they are collected, the analysis giving clues as to the most fruitful area of further data collection and subsequent analysis. On the other hand, when a particular phenomenon is investigated according to a specific predetermined methodology, it might not even be possible to begin the analysis until all the relevant data have been collected.

However, the types of information collection and analysis are not frequently as distinct as they might at first appear. They can, depending on the particular research aims, be closely interlinked. On the one hand, for instance, in exploratory research, the data may be continuously analyzed as they are collected, the analysis giving clues as to the most fruitful area of further data collection and subsequent analysis.

On the other hand, when a particular phenomenon is investigated according to a specific predetermined methodology, it might not even be possible to begin the analysis until all the relevant data have been collected.

When considering what data you might require, consider carefully the sources, the availability and the possible methods of collecting the data. When considering analysis, think about the tools, techniques and resources required. The different research strategies have often distinctly different methods for data collection and analysis.

QUANTITATIVE AND QUALITATIVE RESEARCH

Quite a strong distinction is generally made between quantitative and qualitative research. Not only do the appropriate data have different characteristics, but they also require different techniques for their analysis. Natural science has traditionally concentrated on ‘hard’ quantitative (positivist) analysis, and this was adopted by the human sciences until its shortcomings became evident. As it became increasingly obvious to some researchers that subjective human feelings and emotions were difficult (or impossible) to quantify, qualitative (anti-positivist) analytical methods were evolved, which took more account of the ‘soft’, personal data.

The approach of research that has a description in numbers paradigm is based on positivism. Science is characterized by empirical research; all phenomena can be minimized to empirical indicators which represent the truth. The ontological position of the quantitative paradigm is that there is only one truth, an objective reality that exists independent of human perception. Epistemologically, the investigator and investigated are independent entities. Therefore, the investigator is capable of studying a phenomenon without influencing it or being influenced by it; “inquiry takes place as through a one way mirror” (Guba and Lincoln, 1994: 110). The goal is to measure and analyze causal relationships between variables within a value-free framework (Denzin and Lincoln, 1994). Techniques to ensure this include randomization, blinding, highly structured protocols, and written or orally administered questionnaires with a limited range of predetermined responses. Sample sizes are much larger than those used in qualitative research so that statistical methods to ensure that samples are representative can be used (Carey, 1993).

In contrast, the qualitative paradigm is based on interpretive (Altheide and Johnson, 1994; Kuzel and Like, 1991; Secker et al., 1995) and constructivism (Guba and Lincoln, 1994). Ontologically speaking, there are multiple realities or multiple truths based on one’s construction of reality. Reality is socially constructed (Berger and Luckmann, 1966) and so is constantly changing. On an epistemological level, there is no access to reality independent of our minds, no external referent by which to compare claims of truth (Smith, 1983). The investigator and the object of study are interactively linked so that findings are mutually created within the context of the situation which shapes the inquiry (Guba and Lincoln, 1994; Denzin and Lincoln, 1994). This suggests that reality has no existence prior to the activity of investigation, and reality ceases to exist when we no longer focus on it (Smith, 1983). The emphasis of qualitative research is on process and meanings. Techniques used in qualitative studies include in-depth and focus group interviews and participant observation. Samples are not meant to represent large populations. Rather, small, purposeful samples of articulate respondents are used because they can provide important information, not because they are representative of a larger group (Reid, 1996).

The underlying assumptions of the quantitative and qualitative paradigms result in differences which extend beyond philosophical and methodological debates. The two paradigms have given rise to different journals, different sources of funding, different expertise, and different methods. There are even differences in scientific language used to describe them.

“Validity” to a quantitative researcher would mean that results correspond to how things really are out there in the world, whereas to a qualitative researcher “valid” is a label applied to an interpretation or description with which one agrees (Smith and Heshusius, 1986). Similarly, the phrase “research has shown ” or “the results of research indicate . ” refers to an accurate reflection of reality to the quantitative researcher, but to a qualitative researcher it announces an interpretation that itself becomes reality (Smith and Heshusius, 1986).

The different ideas of the quantitative and qualitative paradigms originated in the positivism-idealism debate of the late 19th century (Smith, 1983). The inherent differences rarely are discussed or acknowledged by those using mixed-method designs. The reasons why may be because the positivist paradigm has become the predominant frame of reference in the physical and social sciences. In addition, research methods are presented as not belonging to or reflecting

Paradigms. Caracelli and Greene (1993) refer to mixed-method designs as those where neither type of method is inherently linked to a particular inquiry paradigm or philosophy. Guba and Lincoln (1989) claim that questions of method are secondary to questions of paradigms. We argue that methods are shaped by and represent paradigms that reflect a particular belief about reality. We also maintain that the assumptions of the qualitative paradigm are based on a worldview not represented by the quantitative paradigm.

Mixed Methods Research

Both quantitative and qualitative research is an approach to inquiry that combines or associates both qualitative and quantitative forms within the same study to address a single research question. It includes philosophical ideas, the benefit of qualitative and quantitative approaches, and the mixing of both approaches inquire.

Thus, it is more than simply collecting and analyzing both kinds of information; it also includes the use of both approaches in tandem so that the overall strength of a study is greater than either qualitative or quantitative research (Creswell & Plano Clark, 2007).

Mixed-method research draws upon both quantitative and qualitative methodological approaches to answer a particular research question. For example, a researcher may start by conducting semi-structured interviews and then use the results from this phase to formulate specific closed-ended survey questions. This is an illustration of what has been termed sequential mixed-methods research (Cresswell, 2003), whereby qualitative data collection and analysis is undertaken first, followed by quantitative data collection and analysis (or vice versa). Important methods create the benefit of an overriding theoretical perspective (for example, participatory research) to guide the entire study design and to drive the choice of methods used. Within the three general mixed-methods research strategies outlined above, a number of different approaches can be taken.

One question concerns whether both approaches are to be given equal weight, or one approach given priority, and this will depend on the individual researcher's goals and orientation as well as the nature of the research question. Another choice concerns the stage, or stages, at which integration of the two approaches will occur.

In relation to this latter issue, Tashakkori and Teddlie (1998) have defined mixed methods research as that which combines quantitative and qualitative methodologies in a specific aspect of a study (for example, data analysis) and contrast this with mixed-model research which combines both approaches at all stages in the research process (for example, conceptualization, data analysis, interpretation). Two concepts closely associated with mixed-methods research are triangulation and pragmatism.

Triangulation refers to a research strategy that involves approaching a research question from two or more angles in order to converge and cross-validate findings from a number of sources (for example, different data sources). A researcher may converge self-report data derived from interviews with observational data. In combining both quantitative and qualitative approaches mixed methods research embodies the notion of triangulation.

Pragmatism is the term given to a particular paradigm (or 'worldview') which has been identified as providing a rationale for mixed methods research. Pragmatism rejects the traditional conception that the paradigms underlying quantitative and qualitative approaches (positivism and constructivism, respectively) are essentially incompatible and in conflict. Instead, pragmatists argue that both quantitative and qualitative approaches have their own distinctive strengths and weaknesses and can be usefully combined to complement one another. Essentially pragmatism advocates using whatever 'works best' in any particular research context, and thus opens the way for mixed-methods approaches. Pragmatism is a form of methodological pluralism.

Evaluation

As with any approach, **mixed-method research** embodies a number of advantages and disadvantages.

Among the advantages is the potential for gaining a fuller, richer and more complete understanding of a research question by combining both quantitative and qualitative perspectives. This potential is one of the key elements emphasized by pragmatists. Secondly, the results from using one approach, strategy or method may help to guide and inform another approach or method, as in the case where the results of a semi-structured interview provide a useful basis for devising more specific questions for a structured survey.

A potential disadvantage of mixed-methods approaches, however, is the often lengthy data collection and analysis phases required (especially in sequential designs) leading to heavy demands on both time and funding resources. A further possible disadvantage is the demand placed on the researcher to be expert in the use of both quantitative and qualitative approaches. Thirdly, the validity of mixed methods research has been called into question in debates over the extent to which the underlying paradigms and methods of quantitative and qualitative research can be seen as compatible. There have been advocates for both the 'compatibility' and the 'incompatibility' theses. While these theoretical debates are ongoing, a number of researchers across a range of disciplines in the social and behavioral sciences are utilizing and further refining mixed methods approaches.

Collecting Data

The world is full of potential data. The core methods for decreasing the delimitation of your information collection are to study an accurate population, i.e. a small section of the subject of your study. There are several things one must consider in selecting a sample.

SAMPLING

When conducting any kind of survey to collect information, or when choosing some particular cases to study in detail, the question inevitably arises: how representative is the information collected of the whole population? When conducting any kind of survey to collect information, or when choosing some particular cases to study in detail, the question inevitably arises: how representative is the information collected of the whole population?

When we talk about population in research, it does not necessarily mean a number of people. 'Population' is a collective term used to describe the total quantity of cases of the type which are the subject of your study. So a population can consist of objects, people or even events, e.g. schools, miners, revolutions.

A sample is a selected number of cases in a population. You will have to devise some way of selecting a sample of the members who you are able to question, and who are a fair representation of all the members of the union. Sampling must be done whenever you can gather information from only a fraction of the population of a group or a phenomenon which you want to study. Ideally, you should try to select a sample which is free from bias. You will see that the type of sample you select will greatly affect the reliability of your subsequent generalizations.

There are basically two types of sampling procedure – random and nonrandom. Random sampling techniques give the most reliable representation of the whole population, while non-random techniques, relying on the judgment of the researcher or on accident, cannot generally be used to make generalizations about the whole population.

RANDOM SAMPLING

The simplest form of random sampling is to represent all the units (sometimes called elements or cases) in a population on slips of paper, put them into a hat, and draw out the slips in a random fashion. However, for a researcher facing the practicalities of a specific research situation, this example is not very useful, so a set of guidelines is called for.

First, a question should be asked about the nature of the population: is it homogeneous or are there distinctly different classes of cases within it? Different sampling techniques are appropriate for each. The next question to ask is: which process of randomization will be used? The following gives a guide to which technique is suited to the different population characteristics.

Simple random sampling is used when the population is uniform or has similar characteristics in all cases, e.g. a production batch of cars of a particular model from which random samples are selected for testing as to their quality.

When the population is not quite as uniform or one-dimensional as a particular model of a car, simple random sampling is not quite as simple as it sounds. The procedure should aim to guarantee that each element (person, group, class, type etc.) has an equal chance of being selected and that every possible combination of the elements also has an equal chance of being selected.

While it is virtually impossible to achieve this in practice, several methods, some using randomly generated numbers, have been devised to produce some form of a fair lottery in which each combination of elements has an equal chance in coming up.

Simple stratified sampling should be used when cases in the population fall into distinctly different categories (strata), e.g. a business whose workforce is divided into the three categories of production, research and management. With the presence of distinctly different strata in a population, in order to achieve simple randomized sampling, an equally sized randomized sample is obtained from each stratum separately to ensure that each is equally represented. The samples are then combined to form the complete sample from the whole population.

Proportional stratified sampling is used when the cases in a population fall into distinctly different categories (strata) of a known proportion of that population, e.g. a university in which the proportions of the students studying arts and sciences is 61% and 39%.

When the proportions of the different strata in a population are known, then each stratum must be represented in the same proportions within the overall sample. In order to achieve proportional randomized sampling, a randomized sample is obtained from each stratum separately, sized according to the known proportion of each stratum in the whole population, and then combined as previously to form the complete sample from the population.

In **cluster sampling**, cases in the population form clusters by sharing one or some characteristics but are otherwise as heterogeneous as possible, e.g. travelers using main railway stations. They are all train travelers, with each cluster experiencing a distinct station, but individuals vary as to age, sex, nationality, wealth, social status etc. Also known as area sampling, cluster sampling is used when the population is large and spread over a large area. Rather than enumerating the whole population, it is divided into segments, and then several segments are chosen at random. Samples are subsequently obtained from each of these segments using one of the above sampling methods.

Systematic sampling is used when the population is very large and of no known characteristics, e.g. the population of a town. Systematic sampling procedures involve the selection of units in a series (for instance, on a list) according to a predetermined cycles.

There are many possible systems. Perhaps the simplest is to choose every nth case on a list, for example, every tenth person in a telephone directory or list of ratepayers. In using this system, it is important to pick the first case randomly, i.e. the first case on the list is not necessarily chosen. The type of list is also significant: not everyone in the town owns a telephone or is a ratepayer.

CONVENIENCE SAMPLING

Although non-random sampling can be useful for certain studies, it provides only a weak basis for generalization. Accidental sampling (or convenience sampling) involves using what is immediately available, e.g. studying the building you happen to be in, examining the work practices in your firm etc.

There are no ways of checking to see if this kind of sample is in any way representative of others of its kind, so the results of the study can be applied only to that sample. Used regularly by reporters interviewing on the streets, quota sampling is an attempt to balance the sample interviewed by selecting responses from equal numbers of different respondents, e.g. equal numbers from different political parties. This is an unregulated form of sampling, as there is no knowledge of whether the respondents are typical of their parties. For example, Labor respondents might just have come from an extreme leftwing rally.

A useful method of getting information from a sample of the population that you think knows most about a subject is theoretical sampling. A study on homelessness could concentrate on questioning people living in the street. This approach is common in qualitative research where statistical inference is not required.

Three other methods can be briefly mentioned: purposive sampling, where the researcher selects what he/she thinks is a ‘typical’ sample; systematic matching sampling, when two groups of very different size are compared by selecting a number from the larger group to match the number and characteristics of the smaller one; and snowball techniques, where you contact a small number of members of the target population and get them to introduce you to others, e.g. of a secret society. Having selected a suitable sampling method, the remaining problem is to determine the sample size.

There is no easy answer to this problem. If the population is very homogeneous, and the study is not very detailed, then a small sample will give a fairly representative view of the whole. In other cases, you should consider the following.

The greater the accuracy required in the true representation of the population, then the larger the sample must be.

The size of the sample also should be in direct relationship to the number of questions asked, the amount of detail required in the analysis of the data, and the number of controls introduced. It is generally accepted that conclusions reached from the study of a large sample are more convincing than those from a small one.

However, the preference for a large sample must be balanced against the practicalities of the research resources, i.e. cost, time and effort.

The amount of variability within the population (technically known as the standard deviation) is another important factor in determining a suitable sample size. Obviously, in order that every sector of a diverse population is adequately represented, a larger sample will be required than if the population were more homogeneous.

If statistical tests are to be used to analyze the data, there are usually minimum sample sizes specified from which any significant results can be obtained. A later part of this chapter deals briefly with statistical methods.

Important Concepts Relating to Research Design

Measurable things towards research

A concept which can take on different quantitative values is called a variable. The term variable refers to a property whereby the members of a group being studied differ from one another. Labels or numbers may be used to describe the way in which one member of a group is the same or different from another.
Some examples are:

• 'Sex' is a variable with two values: The individuals being measured may be either male or female. The values attached for computer processing purposes could be 1 or 2.
• 'Occupational Category' is a variable that may take a range of values depending on the occupational classification scheme that is being used

With variables like 'height', 'age', and 'intelligence', the measurement is carried out by assigning descriptive numerical values. Thus an individual may be 1.5 meters tall, 50 years of age, and have an Intelligence Quotient of 105.
Variables aren't always 'quantitative' or numerical. The variable 'gender' consists of two text values: 'male' and 'female'.
We can, if it is useful, assign quantitative values instead of (or in place of) the text values, but we don't have to assign numbers in order for something to be a variable. It's also important to realize that variables aren't only things that we measure in the traditional sense. For instance, in much social research and in program evaluation, we consider the treatment or program to be made up of one or more variables (i.e., the 'cause' can be considered a variable). An educational program can have varying amounts of 'time on task', 'classroom settings', 'student-teacher ratios', etc. So, even the program can be considered a variable (which can be made up of a number of sub-variables).
An *attribute* is a specific value on a variable. For instance, the variable *sex* or *gender* has two attributes: *Male* and *Female* Or, the variable *agreement* might be defined as having five attributes:

- 1 = strongly disagree
- 2 = disagree
- 3 = neutral
- 4 = agree
- 5 = strongly agree

Another important distinction having to do with the term 'variable' is the distinction between an *independent* and *dependent* variable. This distinction is particularly relevant when you are investigating cause-effect relationships. In fact *the independent variable is what you (or nature) manipulates* -- a treatment or program or cause. The *dependent variable is what is affected by the independent variable* -- your effects or outcomes. For example, if you are studying the effects of a new educational program on student achievement, the program is the independent variable and your measures of achievement are the dependent ones.
In nut shell, there are two traits of variables that should always be achieved. Each variable should be *exhaustive*; it should include all possible answerable responses. For instance, if the variable is "religion" and the only options are "Protestant", "Jewish", and "Muslim", there are quite a few religions I can think of that haven't been included. The list does not exhaust all possibilities.
In addition to being exhaustive, the attributes of a variable should be *mutually exclusive*; no respondent should be able to have two attributes simultaneously. While this might seem obvious, it is often rather tricky in practice. For instance, you might be tempted to represent the variable "Employment Status" with the two attributes "employed" and "unemployed." But these attributes are not necessarily mutually exclusive -- a person who is looking for a second job while employed would be able to check both attributes! But don't we often use questions on surveys that ask the respondent to "check all that apply" and then list a series of categories?

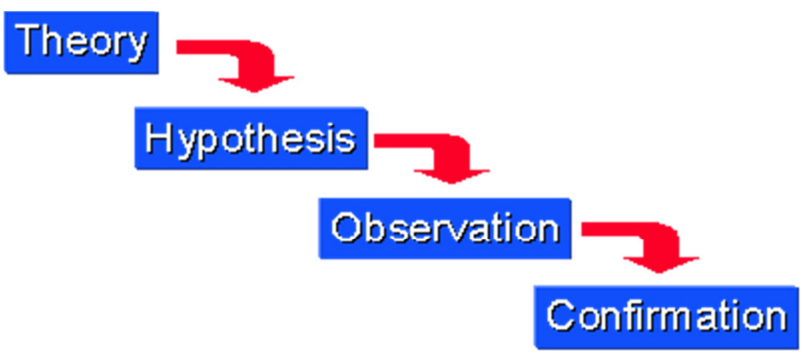
1. **Types of variables:** Variables may be classified according to the type of information which different classifications or measurements provide. There are four main types of variables: nominal, ordinal, interval, and ratio.

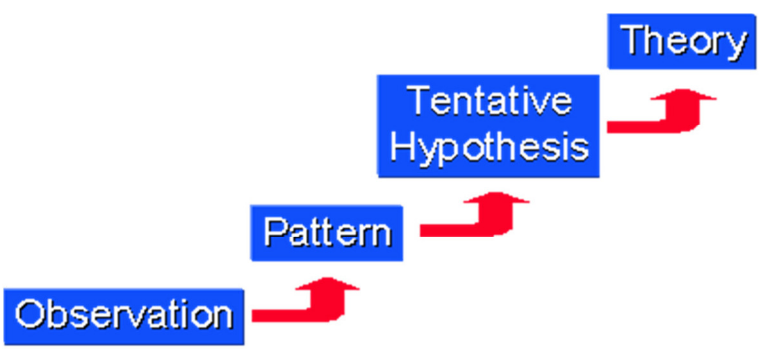
a. **Nominal:** This type of variable permits statements to be made only about equality or difference. Therefore we may say that one individual is the ‘same as’ or’ ‘different from’ another individual. For example, color of hair, religion, country of birth.
b. **Ordinal:** This type of variable permits statements about the rank ordering of the members of a group. Therefore we may make statements about some characteristics of an individual being ‘greater than’ or ‘less than’ other members of a group. For example, physical beauty, agility, happiness.

c. **Interval:** This type of variable permits statements about the rank ordering of individuals. It also permits statements to be made about the ‘size of the intervals’ along the scale which is used to measure the individuals and to compare distances at points along the scale. It is important to note that interval variables do not have true zero points. The numbering of the years in describing dates is an interval scale because the distance between points on the scale is comparable at any point, but the choice of a zero point is a socio-cultural decision.
d. **Ratio:** This type of variable permits all the statements which can be made for the other three types of variables. In addition, a ratio variable has an absolute zero point. This means that a value for this type of variable may be spoken of as ‘double’ or ‘one third of’ another value.

Deductive and Inductive approaches

In logic, we often refer to the two broad methods of reasoning as the ***deductive*** and ***inductive*** approaches.





Deductive reasoning works from the more general to the more specific. Sometimes this is informally called a "top-down" approach. We might begin with thinking up a *theory* about our topic of interest. We then narrow that down into more specific *hypotheses* that we can test. We narrow down even further when we collect *observations* to address the hypotheses. This ultimately leads us to be able to test the hypotheses with specific data -- a *confirmation* (or not) of our original theories.

Inductive reasoning works the other way, moving from specific observations to broader generalizations and theories. Informally, we sometimes call this a "bottom up" approach (please note that it's "bottom up" and *not* "bottom**s** up" which is the kind of thing the bartender says to customers when he's trying to close for the night!). In inductive reasoning, we begin with specific observations and measures, begin to detect patterns and regularities, formulate some tentative hypotheses that we can explore, and finally end up developing some general conclusions or theories.

These two methods of reasoning have a very different "feel" to them when you're conducting research. Inductive reasoning, by its very nature, is more open-ended and exploratory, especially at the beginning. Deductive reasoning is narrower in nature and is concerned with testing or confirming hypotheses.

Even though a particular study may look like it's purely deductive (e.g., an experiment designed to test the hypothesized effects of some treatment on some outcome), most social research involves both inductive and deductive reasoning processes at some time in the project. In fact, it doesn't take a rocket scientist to see that we could assemble the two graphs above into a single circular one that continually cycles from theories down to observations and back up again to theories. Even in the most constrained experiment, the researchers may observe patterns in the data that lead them to develop new theories.

Intended Validity and reliability

Whenever individuals are measured on variables in educational research, there are two main characteristics of the measurement which must be taken into consideration: Validity and Reliability.

Validity

Validity is the most important characteristic to consider when constructing or selecting a test or measurement technique. A valid test or measure is one which measures what it is intended to measure. Validity must always be examined with respect to the use which is to be made of the values obtained from the measurement procedure.

For example, the results from an arithmetic test may have a high degree of validity for indicating skill in numerical calculation, a low degree of validity for indicating general reasoning ability, a moderate degree of validity for predicting success in future mathematics courses, and no validity at all for predicting success in art or music. There are three important types of validity in educational research: content validity, criterion-related validity, and construct validity.

a. Content validity

This type of validity refers to the extent to which a test measures a representative sample of subject-matter content and behavioral content from the syllabus which is being measured. For example, consider a test which has been designed to measure “Competence in Using the English Language”.

In order to examine the content validity of the test one must initially examine the subject-matter knowledge and the behavioral skills which were required to complete the test, and then after this examination compare these to the subject-matter knowledge and behavioral skills which are agreed to comprise correct and effective use of the English language. The test would have high content validity if there was a close match between these two areas.

b. Criterion-related validity

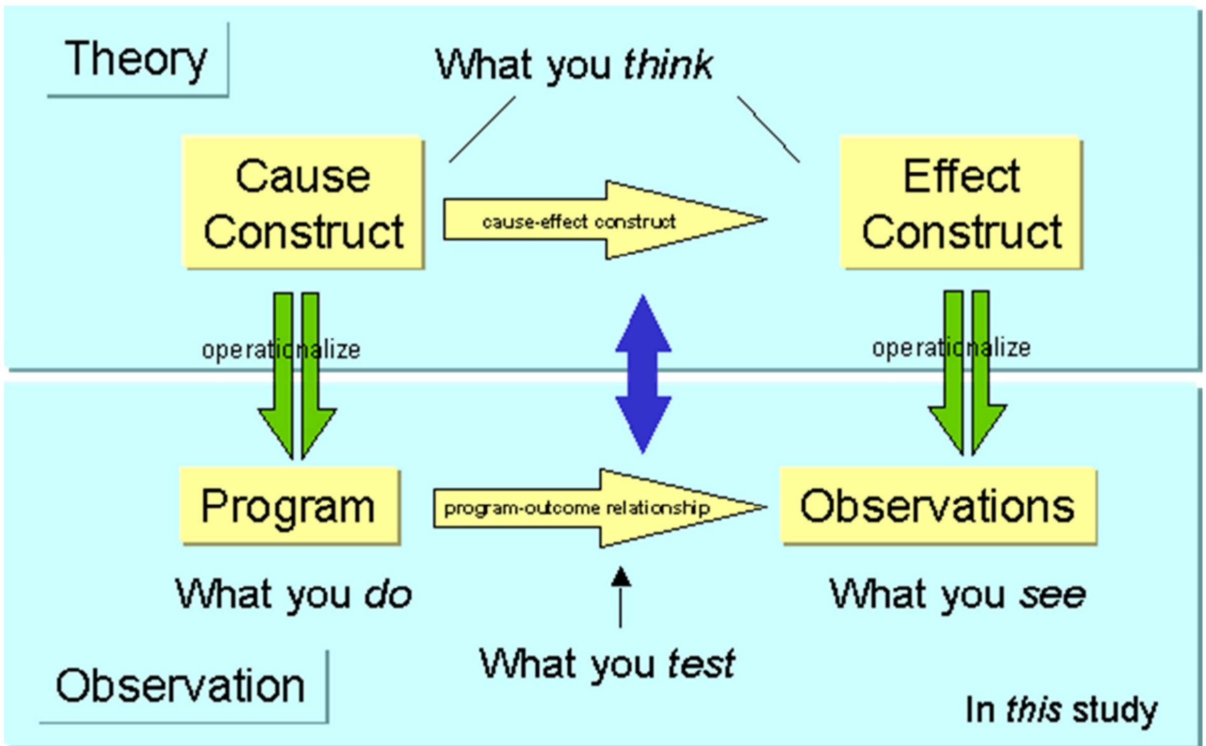
This type of validity refers to the capacity of the test scores to predict future performance or to estimate current performance on some valued measure other than the test itself. For example, ‘Reading Readiness’ scores might be used to predict a student’s future reading achievement, or a test of dictionary skills might be used to estimate a student’s skill in the use of the dictionary (as determined by observation).

In the first example, the interest is in prediction and thus in the relationship between the two measures over an extended period of time. In the second example the interest is in estimating present status and thus in the relationship between the two measures obtained concurrently.

c. Construct validity

This type of validity is concerned with the extent to which test performance can be interpreted in terms of certain psychological constructs. A construct is a psychological quality which is assumed to exist in order to explain some aspect of behavior.

For example, “Reasoning Ability” is a construct. When test scores are interpreted as measures of reasoning ability, the implication is that there is a quality associated with individuals that can be properly called reasoning ability and that it can account to some degree for performance on the test.
We subdivide validity into four types. Each type addresses a specific methodological question. In order to understand the types of validity, you have to know something about how we investigate a research question. Because all four validity types are really only operative when studying causal questions, we will use a causal study to set the context.

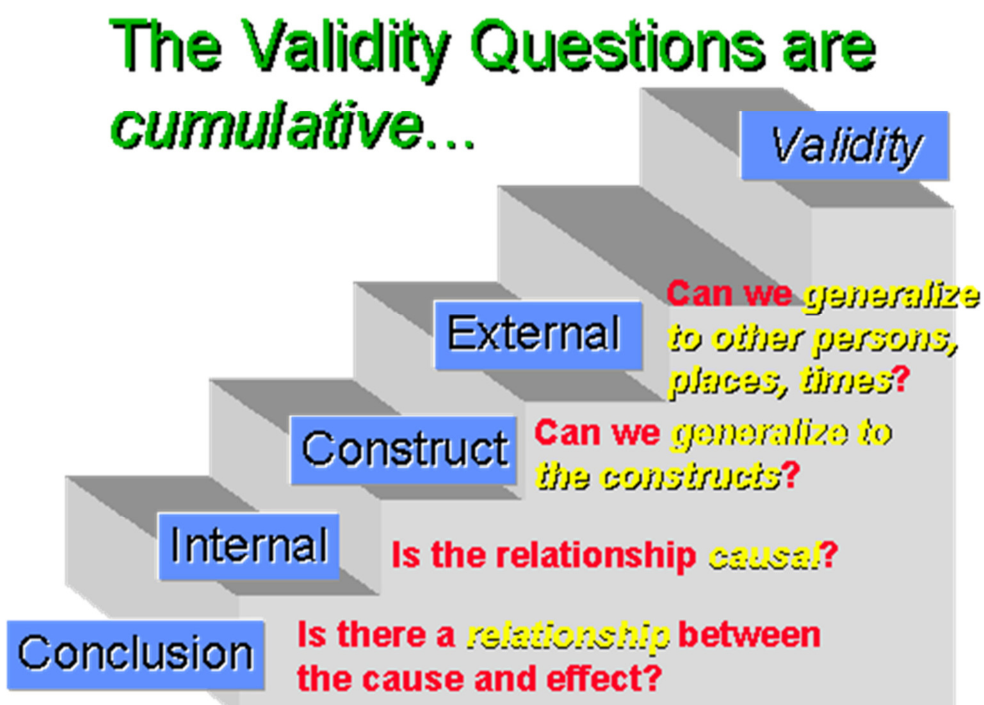


The figure shows that there are really two realms that are involved in research. The first, on the top, is the land of theory. It is what goes on inside our heads as researchers. It is where we keep our theories about how the world operates. The second, on the bottom, is the land of observations. It is the real world into which we translate our ideas -- our programs, treatments, measures and observations. When we conduct research, we are continually flitting back and forth between these two realms, between what we think about the world and what is going on in it. When we are investigating a cause-effect relationship, we have a theory (implicit or otherwise) of what the cause is (***the cause construct***). For instance, if we are testing a new educational program, we have an idea of what it would look like ideally. Similarly, on the effect side, we have an idea of what we are ideally trying to affect and measure (***the effect construct***). But each of these, the cause and the effect, has to be translated into real things, into a program or treatment and a measure or observational method. We use the term ***operationalization*** to describe the act of translating a construct into its manifestation. In effect, we take our idea and describe it as a series of operations or procedures. Now, instead of it only being an idea in our minds, it becomes a public entity that anyone can look at and examine for themselves.
It is one thing, for instance, for you to say that you would like to measure self-esteem (a construct). But when you show a ten-item paper-and-pencil self-esteem measure that you developed for that purpose, others can look at it and understand more clearly what you intend by the term self-esteem.

Imagine that we wish to examine whether use of a World Wide Web (WWW) Virtual Classroom improves student understanding of course material. Assume that we took these two constructs, the cause construct (the WWW site) and the effect (understanding), and operationalized them -- turned them into realities by constructing the WWW site and a measure of knowledge of the course material. Here are the four validity types and the question each addresses:

Validity

In the context of the example we're considering, the question might be worded: in this study, is there a relationship between the WWW site and knowledge of course material? There are several conclusions or inferences we might draw to answer such a question. We could, for example, conclude that there is a relationship. We might conclude that there is a positive relationship. We might infer that there is no relationship.
We can assess the conclusion validity of each of these conclusions or inferences.
Internal Validity: *Assuming that there is a relationship in this study;* is the relationship a ***causal*** one? Just because we find that use of the WWW site and knowledge are correlated, we can't necessarily assume that WWW site use *causes* the knowledge. Both could, for example, be caused by the same factor. For instance, it may be that wealthier students who have greater resources would be more likely to use have access to a WWW site and would excel on objective tests. When we want to make a claim that our program or treatment caused the outcomes in our study, we can consider the internal validity of our causal claim.
Construct Validity: *Assuming that there is a causal relationship in this study*; can we claim that the program reflected well our ***construct*** of the program and that our measure reflected well our idea of the ***construct*** of the measure?
In simpler terms, did we implement the program we intended to implement and did we measure the outcome we wanted to measure? In yet other terms, did we operationalize well the ideas of the cause and the effect? When our research is over, we would like to be able to conclude that we did a credible job of operational zing our constructs -- we can assess the construct validity of this conclusion.
External Validity: *Assuming that there is a causal relationship in this study between the constructs of the cause and the effect*; can we ***generalize*** this effect to other persons, places or times?
Conclusion: In this study, is there a ***relationship*** between the two variables?



Notice how the question that each validity type addresses presupposes an affirmative answer to the previous one. This is what we mean when we say that the validity types build on one another. The figure shows the idea of cumulativeness as a staircase, along with the key question for each validity type.
For any inference or conclusion, there are always possible ***threats to validity*** -- reasons the conclusion or inference

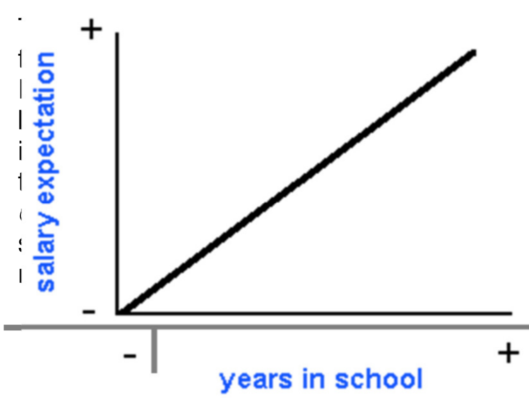
might be wrong. Ideally, one tries to reduce the plausibility of the most likely threats to validity, thereby leaving as most plausible the conclusion reached in the study. For instance, imagine a study examining whether there is a relationship between the amount of training in a specific technology and subsequent rates of use of that technology. Because the interest is in a relationship, it is considered an issue of conclusion validity. Assume that the study is completed and no significant correlation between amount of training and adoption rates is found. On this basis it is *concluded* that there is no relationship between the two. How could this conclusion be wrong -- that is, what are the "threats to validity"? For one, it's possible that there isn't sufficient statistical power to detect a relationship even if it exists. Perhaps the sample size is too small or the measure of amount of training is unreliable. Or maybe assumptions of the correlation test are violated given the variables used. Perhaps there were random irrelevancies in the study setting or random heterogeneity in the respondents that increased the variability in the data and made it harder to see the relationship of interest. The inference that there is no relationship will be stronger -- have greater conclusion validity -- if one can show that these alternative explanations are not credible. The distributions might be examined to see if they conform to assumptions of the statistical test, or analyses conducted to determine whether there is sufficient statistical power.
The theory is general in scope and applicability, well-articulated in its philosophical suppositions, and virtually impossible to explain adequately in a few minutes. As a framework for judging the quality of evaluations it is indispensable and well worth understanding.

Types of Relationships

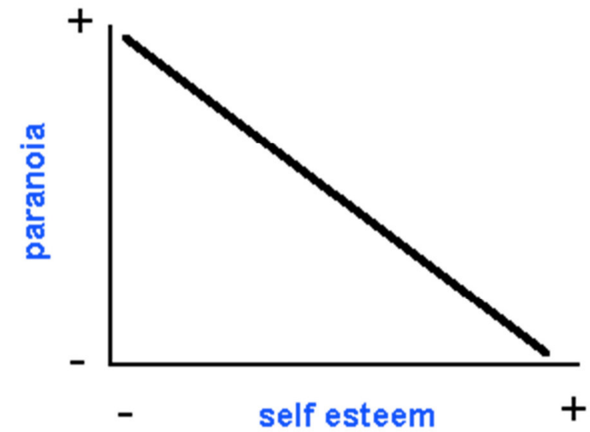
A relationship refers to the *correspondence* between two variables. When we talk about types of relationships, we can mean that in at least two ways: the *nature* of the relationship or the *pattern* of it.

The Nature of a Relationship

While all relationships tell about the correspondence between two variables, there is a special type of relationship that holds that the two variables are not only in correspondence, but that one *causes* the other. This is the key distinction between a simple ***correlation relationship*** and a ***causal relationship***. A co relational relationship simply says that two things perform in a synchronized manner. For instance, there has often been talk of a relationship between ability in math and proficiency in music. In general people who are good in one may have a greater tendency to be good in the other; those who are poor in one may also tend to be poor in the other. If this relationship is true, then we can say that the two variables are correlated. But knowing that two variables are correlated does not tell us whether one *causes* the other. We know, for instance, that there is a correlation between the number of roads built in Europe and the number of children born in the United States. Does that mean that if we want fewer children in the U.S., we should stop building so many roads in Europe? Or, does it mean that if we don't have enough roads in Europe, we should encourage U.S. citizens to have more babies? Of course not (At least, I hope not). While there is a relationship between the number of roads built and the number of babies, we don't believe that the relationship is a *causal* one.



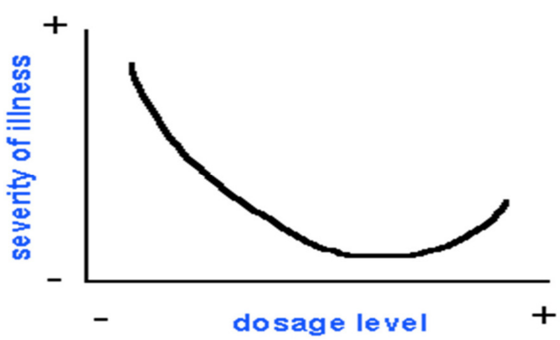
termed the ***third variable problem***. In this example, it may be that there is a
ing of roads and the birthrate, which is causing the correlation we observe.
onomy is responsible for both. When the economy is good more roads are
in the U.S. The key lesson here is that you have to be careful when you
lation between the number of hours students use the computer to study and
uter users getting higher grades), you *cannot* assume that the relationship is
s. In this case, the third variable might be socioeconomic status -- richer
ir disposal tend to both use computers and do better in their grades. It's the
ot computer use that causes the change in the grade point average.



Patterns of Relationships

We have several terms to describe the major different types of patterns one might find in a relationship. First, there is the case of ***no relationship*** at all. If you know the values on one variable, you don't know anything about the values on the other. For instance, I suspect that there is no relationship between the length of the lifeline on your hand and your grade point average. If I know your GPA, I don't have any idea how long your lifeline is.

Then, we have the ***positive relationship***. In a positive relationship, high values on one variable are associated with high values on the other and low values on one are associated with low values on the other. In this example, we assume an idealized positive relationship between years of education and the salary one might expect to be making.



On the other hand a ***negative*** relationship implies that high values on one variable are associated with low values on the other. This is also sometimes termed an ***inverse*** relationship. Here, we show an idealized negative relationship between a measure of self-esteem and a measure of paranoia in psychiatric patients.
These are the simplest types of relationships we might typically estimate in research. But the pattern of a relationship can be more complex than this. For instance, the figure on the left shows a relationship those changes over the range of both variables, a curvilinear relationship. In this example, the horizontal axis represents dosage of a drug for an illness and the vertical axis represents a severity of illness measure. As dosage rises, severity of illness goes down. But at some point, the patient begins to experience negative side effects associated with too high a dosage, and the severity of illness begins to increase again.

Logic is an error in reasoning, usually based on mistaken assumptions. Researchers are very familiar with all the ways they could go wrong, with the fallacies they are susceptible to. Two of the most important ones are discussed here.

The ***ecological fallacy*** occurs when you make conclusions about individuals based only on analyses of group data. For instance, assume that you measured the math scores of a particular classroom and found that they had the highest average score in the district. Later (probably at the mall) you run into one of the kids from that class and you think to yourself "she must be a math whiz." Aha! Fallacy! Just because she comes from the class with the highest *average* doesn't mean that she is automatically a high-scorer in math. She could be the lowest math scorer in a class that otherwise consists of math geniuses!

An ***exception fallacy*** is sort of the reverse of the ecological fallacy. It occurs when you reach a group conclusion on the basis of exceptional cases. This is the kind of fallacious reasoning that is at the core of a lot of sexism and racism. The stereotype is of the guy who sees a woman make a driving error and concludes that "women are terrible drivers." Wrong! Fallacy!

Both of these fallacies point to some of the traps that exist in both research and everyday reasoning. They also point out how important it is that we do research. We need to determine empirically how individuals perform (not just rely on group averages). Similarly, we need to look at whether there are correlations between certain behaviors and certain groups (you might look at the whole controversy around the book *The Bell Curve* as an attempt to examine whether the supposed relationship between race and IQ is real or a fallacy.

Reliability

Reliability refers to the degree to which a measuring procedure gives consistent results. That is, a reliable test is a test which would provide a consistent set of scores for a group of individuals if it was administered independently on several occasions.

Reliability is a necessary but not sufficient condition for validity. A test which provides totally inconsistent results cannot possibly provide accurate information about the behavior being measured. Thus low reliability can be expected to restrict the degree of validity that is obtained, but high reliability provides no guarantee that a satisfactory degree of validity will be present.

Note that reliability refers to the nature of the test scores and not to the test itself. Any particular test may have a number of different reliabilities, depending on the group involved and the situation in which it is used. The assessment of reliability is measured by the ‘Reliability Coefficient’ (for groups of individuals) or the “Standard Error of Measurement” (for individuals).

The ethics research

We are going through a time of profound change in our understanding of the ethics of applied social research. From the time immediately after World War II until the early 1990s, there was a gradually developing consensus about the key ethical principles that should underlie the research endeavor. The Nuremberg War Crimes Trial following World War II brought to public view the ways German scientists had used captive human subjects as subjects in oftentimes gruesome experiments. In the 1950s and 1960s, the Tuskegee Syphilis Study involved the withholding of known effective treatment for syphilis from African- American participants who were infected. Events like these forced the reexamination of ethical standards and the gradual development of a consensus that potential human subjects needed to be protected from being used as 'guinea pigs' in scientific research.

By the 1990s, the dynamics of the situation changed. Cancer patients and persons with AIDS fought publicly with the medical research establishment about the long time.

In many cases, it is the ethical assumptions of the previous thirty years that drive this 'go-slow' mentality. After all, we would rather risk denying treatment for a while until we achieve enough confidence in a treatment, rather than run the risk of harming innocent people (as in the Nuremberg and Tuskegee events). But now, those who were threatened with fatal illness were saying to the research establishment that they *wanted* to be test subjects, even under experimental conditions of considerable risk. You had several very vocal and articulate patient groups who wanted to be experimented on coming up against an ethical review system that was designed to protect them from being experimented on.

Although the last few years in the ethics of research have been tumultuous ones, it is beginning to appear that a new consensus is evolving that involves the stakeholder groups most affected by a problem participating more actively in the formulation of guidelines for research. While it's not entirely clear, at present, what the new consensus will be, it is almost certain that it will not fall at either extreme: protecting against human experimentation at all costs **vs.** allowing anyone who is willing to be experimented on.

Ethical Issues

There are a number of key phrases that describe the system of ethical protections that the contemporary social and medical research establishments have created to try to protect better the rights of their research participants. The principle of ***voluntary participation*** requires that people not be coerced into participating in research. This is especially relevant where researchers had previously relied on 'captive audiences' for their subjects -- prisons, universities, and places like that. Closely related to the notion of voluntary participation is the requirement of ***informed consent***.

Essentially, this means that prospective research participants must be fully informed about the procedures and risks involved in research and must give their consent to participate. Ethical standards also require that researchers not put participants in a situation where they might be at ***risk of harm*** as a result of their participation.

Harm can be defined as both physical and psychological. There are two standards that are applied in order to help protect the privacy of research participants. Almost all research guarantees the participants' ***confidentiality*** -- they are assured that identifying information will not be made available to anyone who is not directly involved in the study. The stricter standard is the principle of ***anonymity*** which essentially means that the participant will remain anonymous throughout the study -- even to the researchers themselves. Clearly, the anonymity standard is a stronger guarantee of privacy, but it is sometimes difficult to accomplish, especially in situations where participants have to be measured at multiple time points (e.g., a pre-post study). Increasingly, researchers have had to deal with the ethical issue of a person's ***right to service***. Good research practice often requires the use of a no-treatment control group -- a group of participants who do *not* get the treatment or program that is being studied. But when that treatment or program may have beneficial effects, persons assigned to the no-treatment control may feel their rights to equal access to services are being curtailed.

Even when clear ethical standards and principles exist, there will be times when the need to do accurate research runs up against the rights of potential participants. No set of standards can possibly anticipate every ethical circumstance. Furthermore, there needs to be a procedure that assures that researchers will consider all relevant ethical issues in formulating research plans. To address such needs most institutions and organizations have formulated an ***Institutional Review Board (IRB)***, a panel of persons who reviews grant proposals with respect to ethical implications and decides whether additional actions need to be taken to assure the safety and rights of participants. By reviewing proposals for research, IRBs also help to protect both the organization and the researcher against potential legal implications of neglecting to address important ethical issues of participants.

Extra information

Techniques of data gathering

Sampling is the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen.
Let's begin by covering some of the key terms in sampling like "population" and "sampling frame." Then, because some types of sampling rely upon quantitative models, we'll talk about some of the statistical terms used in sampling. Finally, we'll discuss the major distinction between probability and Nonprobability sampling methods and work through the major types in each.

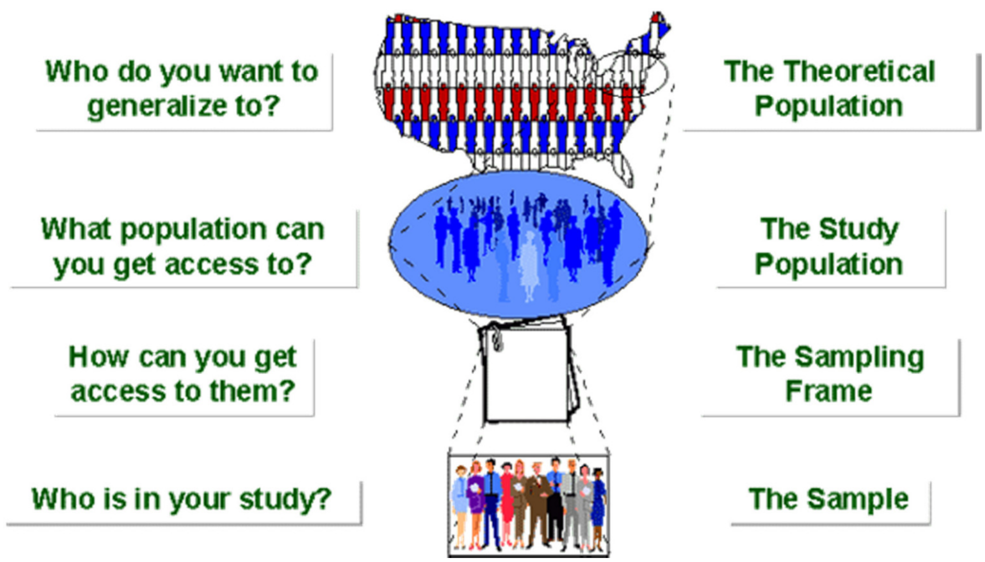
Sampling Terminology

As with anything else in life you have to learn the language of an area if you're going to ever hope to use it. Here, I want to introduce several different terms for the major groups that are involved in a sampling process and the role that each group plays in the logic of sampling.

The major question that motivates sampling in the first place is: "Who do you want to generalize to?" Or should it be: "To whom do you want to generalize?" In most social research we are interested in more than just the people who directly participate in our study. We would like to be able to talk in general terms and not be confined only to the people who are in our study. Now, there are times when we aren't very concerned about generalizing. Maybe we're just evaluating a program in a local agency and we don't care whether the program would work with other people in other places and at other times. In that case, sampling and generalizing might not be of interest. In other cases, we would really like to be able to generalize almost universally. When psychologists do research, they are often interested in developing theories that would hold for all humans. But in most applied social research, we are interested in generalizing to specific groups. The group you wish to generalize to is often called the **population** in your study. This is the group you would like to sample from because this is the group you are interested in generalizing to.

Let's imagine that you wish to generalize to urban homeless males between the ages of 30 and 50 in the United States. If that is the population of interest, you are likely to have a very hard time developing a reasonable sampling plan. You are probably not going to find an accurate listing of this population, and even if you did, you would almost certainly not be able to mount a national sample across hundreds of urban areas. So we probably should make a distinction between the population you would like to generalize to, and the population that will be accessible to you.

We'll call the former the **theoretical population** and the latter the **accessible population**. In this example, the accessible population might be homeless males between the ages of 30 and 50 in six selected urban areas across the U.S.



Once you've identified the theoretical and accessible populations, you have to do one more thing before you can actually draw a sample -- you have to get a list of the members of the accessible population. (Or, you have to spell out in detail how you will contact them to assure representativeness). The listing of the accessible population from which you'll draw your sample is called the **sampling frame**. If you were doing a phone survey and selecting names from the telephone book, the book would be your sampling frame.

That wouldn't be a great way to sample because significant sub portions of the population either don't have a phone.

Notice that in this case, you might identify the area code and all three-digit prefixes within that area code and draw a sample simply by randomly dialing numbers (cleverly known as *random-digit-dialing*). In this case, the sampling frame is not a list *per se*, but is rather a procedure that you follow as the actual basis for sampling. Finally, you actually draw your sample (using one of the many sampling procedures).

The **sample** is the group of people who you select to be in your study. Notice that I didn't say that the sample was the group of people who are actually *in* your study. You may not be able to contact or recruit all of the people you actually sample, or some could drop out over the course of the study. The group that actually completes your study is a subsample of the sample -- it doesn't include correspondents or dropouts.

On the spot of this idea, you should appreciate that sampling is a difficult multi-step process and that there are lots of places you can go wrong. In fact, as we move from each step to the next in identifying a sample, there is the possibility of introducing systematic error or **bias**. For instance, even if you are able to identify perfectly the population of interest, you may not have access to all of them. And even if you do, you may not have a complete and accurate enumeration or sampling frame from which to select. And, even if you do, you may not draw the sample correctly or accurately. And, even if you do, they may not all come and they may not all stay. Depressed yet? This is a very difficult business indeed. At times like this I'm reminded of what Donald Campbell used to say (I'll paraphrase here): "Cousins to the amoeba, it's amazing that we know anything at all!"

Statistical Terms in Sampling

Let's begin by defining some very simple terms that are relevant here. First, let's look at the results of our sampling efforts. When we sample, the units that we sample -- usually people -- supply us with one or more responses. In this sense, a **response** is a specific measurement value that a sampling unit supplies. In the figure, the person is responding to a survey instrument and gives a response of '4'. When we look across the responses that we get for our entire sample, we use a **statistic**. There are a wide variety of statistics we can use -- mean, median, mode, and so on. In this example, we see that the mean or average for the sample is 3.75. But the reason we sample is so that we might get an estimate for the population we sampled from. If we could, we would much prefer to measure the entire population. If you measure the entire population and calculate a value like a mean or average, we don't refer to this as a statistic, we call it a **parameter** of the population.

The Sampling Distribution

So how do we get from our sample statistic to an estimate of the population parameter? A crucial midway concept you need to understand is the **sampling distribution**.

In order to understand it, you have to be able and willing to do a thought experiment. Imagine that instead of just taking a single sample like we do in a typical study, you took three independent samples of the same population. And furthermore, imagine that for each of your three samples, you collected a single response and computed a single statistic, say, the mean of the response. Even though all three samples came from the same population, you wouldn't expect to get the exact same statistic from each. They would differ slightly just due to the random "luck of the draw" or to the natural fluctuations or vagaries of drawing a sample. But you would expect that all three samples would yield a similar statistical estimate because they were drawn from the same population. Now, for the leap of imagination! Imagine that you did an *infinite* number of samples from the same population and computed the average for each one. If you plotted them on a histogram or bar graph you should find that most of them converge on the same central value and that you get fewer and fewer samples that have averages farther away up or down from that central value.

In other words, the bar graph would be well described by the *bell curve* shape that is an indication of a "normal" distribution in statistics. The distribution of an infinite number of samples of the same size as the sample in your study is known as the **sampling distribution**. We don't ever actually construct a sampling distribution. Why not? You're not paying attention! Because to construct it we would have to take an *infinite* number of samples and at least the last time I checked, on this planet infinite is not a number we know how to reach. So why do we even talk about a sampling distribution? Now that's a good question!

Because we need to realize that our sample is just one of a potentially infinite number of samples that we could have taken. When we keep the sampling distribution in mind, we realize that while the statistic we got from our sample is probably near the center of the sampling distribution (because most of the samples would be there) we could have gotten one of the extreme samples just by the luck of the draw.

But what is the standard deviation of the sampling distribution (OK, never had statistics? There are any numbers of places on the web where you can learn about them or even just brush up if you've gotten rusty.

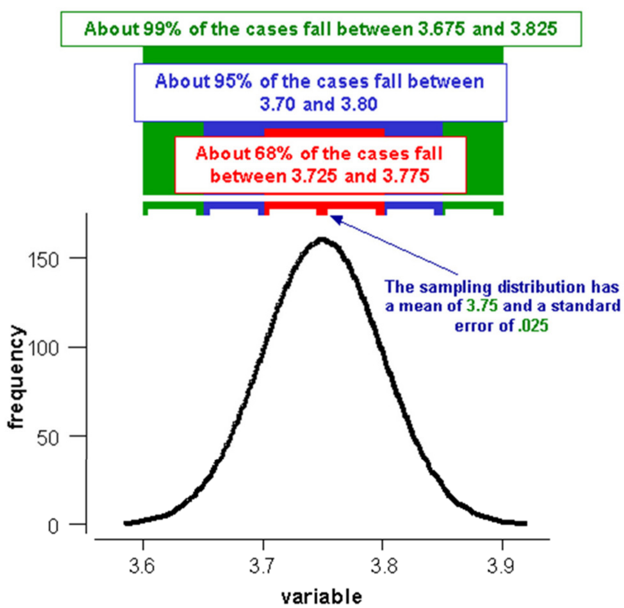
This isn't one of them. I'm going to assume that you at least know what a standard deviation is or that you're capable of finding out relatively quickly). The standard deviation of the sampling distribution tells us something about how different samples would be distributed. In statistics it is referred to as the **standard error** (so we can keep it separate in our minds from standard deviations. Getting confused? Go get a cup of coffee and come back in ten minutes...OK, let's try once more... A standard deviation is the spread of the scores around the average in a single sample. The standard error is the spread of the averages around the average of averages in a sampling distribution. Got it?)

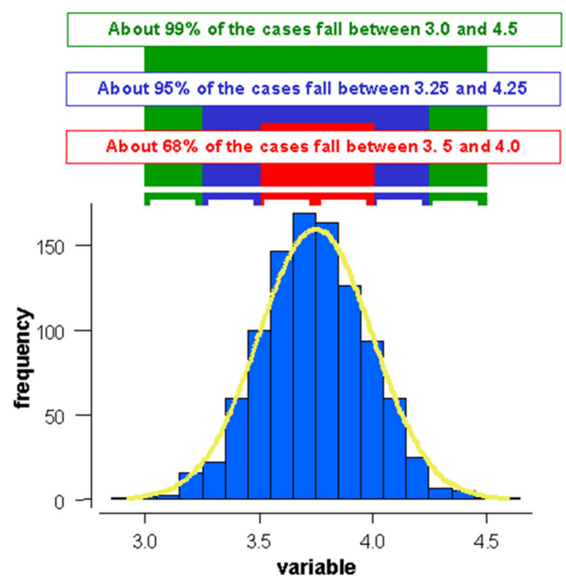
Sampling Error

In sampling contexts, the standard error is called **sampling error**. Sampling error gives us some idea of the precision of our statistical estimate.
A low sampling error means that we had relatively less variability or range in the sampling distribution. But here we go again -- we never actually see the sampling distribution! So how do we calculate sampling error? The standard error is also related to the sample size. The greater your sample size, the smaller the standard error. Why? Because the greater the sample size, the closer your sample is to the actual population itself. If you take a sample that consists of the entire population you actually have no sampling error because you don't have a sample, you have the entire population. In that case, the mean you estimate is the parameter.

The 68, 95, 99 Percent Rule

You've probably heard this one before, but it's so important that it's always worth repeating... There is a general rule that applies whenever we have a normal or bell-shaped distribution. Start with the average -- the center of the distribution. If you go up and down (i.e., left and right) one standard unit, you will include approximately 68% of the cases in the distribution (i.e., 68% of the area under the curve). If you go up and down two standard units, you will include approximately 95% of the cases. And if you go plus-and-minus three standard units, you will include about 99% of the cases.
Notice that I didn't specify in the previous few sentences whether I was talking about standard deviation units or standard error units. That's because the same rule holds for both types of distributions (i.e., the raw data and sampling distributions). For instance, in the figure, the mean of the distribution is 3.75 and the standard unit is .25 (If this was a distribution of raw data, we would be talking in standard deviation units. If it's a sampling distribution, we'd be talking in standard error units). If we go up and down one standard unit





Now, if we have the mean of the sampling distribution (or set it to the mean from our sample) and we have an estimate of the standard error (we calculate that from our sample) then we have the two key ingredients that we need for our sampling distribution in order to estimate confidence intervals for the population parameter.
Perhaps an example will help. Let's assume we did a study and drew a single sample from the population. Furthermore, let's assume that the average for the sample was 3.75 and the standard deviation was .25. This is the raw data distribution depicted above. Now, what would the sampling distribution be in this case? Well, we don't actually

construct it (because we would need to take an infinite number of samples) but we *can* estimate it. For starters, we assume that the mean of the sampling distribution is the mean of the sample, which is 3.75. Then, we calculate the standard error. To do this, we use the standard deviation for our sample and the sample size (in this case N=100) and we come up with a standard error of .025 (just trust me on this). Now we have everything we need to estimate a confidence interval for the population parameter. We would estimate that the probability is 68% that the true parameter value falls between 3.725 and 3.775 (i.e., 3.75 plus and minus .025); that the 95% confidence interval is 3.700 to 3.800; and that we can say with 99% confidence that the population value is between 3.675 and 3.825. The real value (in this fictitious example) was 3.72 and so we have correctly estimated that value with our sample.

Probability Sampling

A **probability sampling** method is any method of sampling that utilizes some form of *random selection*.
In order to have a random selection method, you must set up some process or procedure that assures that the different units in your population have equal probabilities of being chosen. Humans have long practiced various forms of random selection, such as picking a name out of a hat, or choosing the short straw. These days, we tend to use computers as the mechanism for generating random numbers as the basis for random selection.

Some Definitions

Before I can explain the various probability methods we have to define some basic terms. These are:

- **N** = the number of cases in the sampling frame
- **n** = the number of cases in the sample
- $_{N}C_{n}$ = the number of combinations (subsets) of n from N
- **f** = n/N = the sampling fraction

That's it. With those terms defined we can begin to define the different probability sampling methods.

Simple Random Sampling

The simplest form of random sampling is called **simple random sampling**. Here's the quick description of simple random sampling:

- **Objective**: To select *n* units out of *N* such that each $_{N}C_{n}$ has an equal chance of being selected.
- **Procedure**: Use a table of random numbers, a computer random number generator, or a mechanical device to select the sample.

Let's see if we can make it a little more real. How do we select a simple random sample? Let's assume that we are doing some research with a small service agency that wishes to assess client's views of quality of service over the past year.

First, we have to get the sampling frame organized. To accomplish this, we'll go through agency records to identify every client over the past 12 months. If we're lucky, the agency has good accurate computerized records and can quickly produce such a list. Then, we have to actually draw the sample. Decide on the number of clients you would like to have in the final sample. For the sake of the example, let's say you want to select 100 clients to survey and that there were 1000 clients over the past 12 months.

Then, the sampling fraction is f = n/N = 100/1000 = .10 or 10%. Now, to actually draw the sample, you have several options. You could print off the list of 1000 clients, tear then into separate strips, put the strips in a hat, mix them up real good, close your eyes and pull out the first 100. But this mechanical procedure would be tedious and the quality of the sample would depend on how thoroughly you mixed them up and how randomly you reached in. You would need three Sets of balls numbered 0 to 9, one set for each of the digits from 000 to 999 (if we select 000 we'll call that 1000). Number the list of names from 1 to 1000 and then use the ball machine to select the three digits that selects each person. The obvious disadvantage here is that you need to get the ball machines. (Where do they make those things, anyway? Is there a ball machine industry?).

Neither of these mechanical procedures is very feasible and, with the development of inexpensive computers there is a much easier way. Here's a simple procedure that's especially useful if you have the names of the clients already on the computer. Many computer programs can generate a series of random numbers. Let's assume you can copy and paste the list of client names into a column in an EXCEL spreadsheet. Then, in the column right next to it paste the function =RAND (which is EXCEL's way of putting a random number between 0 and 1 in the cells. Then, sort both columns -- the list of names and the random number -- by the random numbers. This rearranges the list in random order from the lowest to the highest random number. Then, all you have to do is take the first hundred names in this sorted list. You could probably accomplish the whole thing in under a minute.

Simple random sampling is simple to accomplish and is easy to explain to others. Because simple random sampling is a fair way to select a sample, it is reasonable to generalize the results from the sample back to the population. Simple random sampling is not the most statistically efficient method of sampling and you may, just because of the luck of the draw, not get good representation of subgroups in a population. To deal with these issues, we have to turn to other sampling methods.

Stratified Random Sampling

Stratified Random Sampling, also sometimes called *proportional* or *quota* random sampling, involves dividing your population into homogeneous subgroups and then taking a simple random sample in each subgroup. In more formal terms:

Objective: Divide the population into non-overlapping groups (i.e., *strata*) N_1, N_2, N_3, N_i, such that $N_1 + N_2 + N_3 + ... + N_i = N$. Then do a simple random sample of $f = n/N$ in each strata.

There are several major reasons why you might prefer stratified sampling over simple random sampling. First, it assures that you will be able to represent not only the overall population, but also key subgroups of the population, especially small minority groups. If you want to be able to talk about subgroups, this may be the only way to effectively assure you'll be able to. If the subgroup is extremely small, you can use different sampling fractions (f) within the different strata to randomly over-sample the small group (although you'll then have to weigh the within-group estimates using the sampling fraction whenever you want overall population estimates). When we use the same sampling fraction within strata we are conducting *proportionate* stratified random sampling. When we use different sampling fractions in the strata, we call this *disproportionate* stratified random sampling. Second, stratified random sampling will generally have more statistical precision than simple random sampling. This will only be true if the strata or groups are homogeneous. If they are, we expect that the variability within-groups are lower than the variability for the population as a whole. Stratified sampling capitalizes on that fact.

For example, let's say that the population of clients for our agency can be divided into three groups: Caucasian, African-American and Hispanic-American.

Furthermore, let's assume that both the African-Americans and Hispanic-Americans are relatively small minorities of the clientele (10% and 5% respectively). If we just did a simple random sample of n=100 with a sampling fraction of 10%, we would expect by chance alone that we would only get 10 and 5 persons from each of our two smaller groups. And, by chance, we could get fewer than that! If we stratify, we can do better. First, let's determine how many people we want to have in each group.

Let's say we still want to take a sample of 100 from the population of 1000 clients over the past year. But we think that in order to say anything about subgroups we will need at least 25 cases in each group. So, let's sample 50 Caucasians, 25 African-Americans, and 25 Hispanic-Americans. We know that 10% of the population, or 100 clients, are African-American. If we randomly sample 25 of these, we have a within-stratum sampling fraction of 25/100 = 25%. Similarly, we know that 5% or 50 clients are Hispanic-American. So our within-stratum sampling fraction will be 25/50 = 50%. Finally, by subtraction we know that there are 850 Caucasian clients. Our within-stratum sampling fraction for them is 50/850 = about 5.88%. Because the groups are more homogeneous within-group than across the population as a whole, we can expect greater statistical precision (less variance). And, because we stratified, we know we will have enough cases from each group to make meaningful subgroup inferences.

Systematic Random Sampling

Here are the steps you need to follow in order to achieve a **systematic random sample**:

- number the units in the population from 1 to N
- decide on the n (sample size) that you want or need
- k = N/n = the interval size
- randomly select an integer between 1 to k
- then take every kth unit

All of this will be much clearer with an example. Let's assume that we have a population that only has N=100 people in it and that you want to take a sample of n=20. To use systematic sampling, the population must be listed in a random order.

The sampling fraction would be f = 20/100 = 20%. In this case, the interval size, k, is equal to N/n = 100/20 = 5. Now, select a random integer from 1 to 5. In our example, imagine that you chose 4. Now, to select the sample, start with the 4th unit in the list and take every k-th unit (every 5th, because k=5). You would be sampling units 4, 9, 14, 19, and so on to 100 and you would wind up with 20 units in your sample.

For this to work, it is essential that the units in the population are randomly ordered, at least with respect to the characteristics you are measuring. Why would you ever want to use systematic random sampling? For one thing, it is fairly easy to do. You only have to select a single random number to start things off. It may also be more precise than simple random sampling. Finally, in some situations there is simply no easier way to do random sampling. For instance, I once had to do a study that involved sampling from all the books in a library. Once selected, I would have to go to the shelf, locate the book, and record when it last circulated. I knew that I had a fairly good sampling frame in the form of the shelf list (which is a card catalog where the entries are arranged in the order they occur on the shelf). To do a simple random sample, I could have estimated the total number of books and generated random numbers to draw the sample; but how would I find book #74,329 easily if that is the number I selected? I couldn't very well count the cards until I came to 74,329! Stratifying wouldn't solve that problem either. For instance, I could have stratified by card catalog drawer and drawn a simple random sample within each drawer. But I'd still be stuck counting cards. Instead, I did a systematic random sample. I estimated the number of books in the entire collection. Let's imagine it was 100,000. I decided that I wanted to take a sample of 1000 for a sampling fraction of 1000/100,000 = 1%. To get the sampling interval k, I divided N/n = 100,000/1000 = 100. Then I selected a random integer between 1 and 100. Let's say I got 57. Next I did a little side study to determine how thick a thousand cards are in the card catalog (taking into account the varying ages of the cards). Let's say that on average I found that two cards that were separated by 100 cards were about .75 inches apart in the catalog drawer. That information gave me everything I needed to draw the sample. I counted to the 57th by hand and recorded the book information. Then, I took a compass. (Remember those from your high-school math class?

They're the funny little metal instruments with a sharp pin on one end and a pencil on the other that you used to draw circles in geometry class.) Then I set the compass at .75", stuck the pin end in at the 57th card and pointed with the pencil end to the next card (approximately 100 books away).

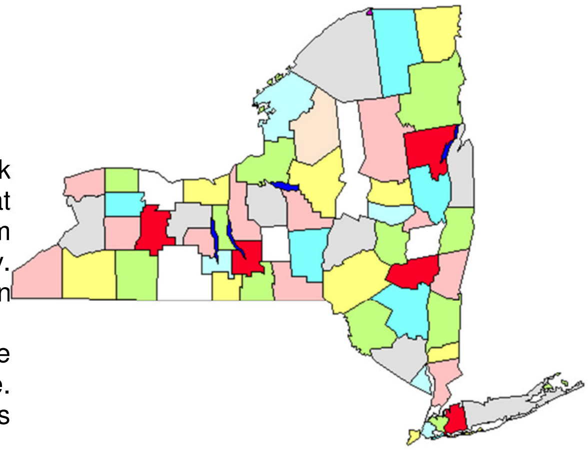
In this way, I approximated selecting the 157th, 257th, 357th, and so on. I was able to accomplish the entire selection procedure in very little time using this systematic random sampling approach. I'd probably still be there counting cards if I'd tried another random sampling method. (Okay, so I have no life. I got compensated nicely, I don't mind saying, for coming up with this scheme.)

Cluster (Area) Random Sampling

The problem with random sampling methods when we have to sample a population that's disbursed across a wide geographic region is that you will have to cover a lot of ground geographically in order to get to each of the units you sampled. Imagine taking a simple random sample of all the residents of New York State in order to conduct personal interviews. By the luck of the draw you will wind up with respondents who come from all over the state. Your interviewers are going to have a lot of traveling to do. It is for precisely this problem that **cluster or area random sampling** was invented.

In cluster sampling, we follow these steps:

- divide population into clusters (usually along geographic boundaries)
- randomly sample clusters
- measure all units within sampled clusters



For instance, in the figure we see a map of the counties in New York State. Let's say that we have to do a survey of town governments that will require us going to the towns personally. If we do simple random samples state-wide we'll have to cover the entire state geographically. Instead, we decide to do a cluster sampling of five counties (marked in red in the figure).

Once these are selected, we go to *every* town government in the five areas. Clearly this strategy will help us to economize on our mileage. Cluster or area sampling, then, is useful in situations like this, and is done primarily for efficiency of administration. Note also, that we probably don't have to worry about using this approach if we are conducting a mail or telephone survey because it doesn't matter as much (or cost more or raise inefficiency) where we call or send letters to.

Multi-Stage Sampling

The four methods we've covered so far -- simple, stratified, and systematic and cluster -- are the simplest random sampling strategies. In most real applied social research, we would use sampling methods that are considerably more complex than these simple variations. The most important principle here is that we can combine the simple methods described earlier in a variety of useful ways that help us address our sampling needs in the most efficient and effective manner possible. When we combine sampling methods, we call this **multi-stage sampling**.

For example, consider the idea of sampling New York State residents for face-to-face interviews. Clearly we would want to do some type of cluster sampling as the first stage of the process. We might sample townships or census tracts throughout the state. But in cluster sampling we would then go on to measure everyone in the clusters we select. Even if we are sampling census tracts we may not be able to measure *everyone* who is in the census tract. So, we might set up a stratified sampling process within the clusters. In this case, we would have a two-stage sampling process with stratified samples within cluster samples. Or, consider the problem of sampling students in grade schools. We might begin with a national sample of school districts stratified by economics and educational level. Within selected districts, we might do a simple random sample of schools. Within schools, we might do a simple random sample of classes or grades. And, within classes, we might even do a simple random sample of students. In this case, we have three or four stages in the sampling process and we use both stratified and simple random sampling. By combining different sampling methods we are able to achieve a rich variety of probabilistic sampling methods that can be used in a wide range of social research contexts.

Non probability Sampling

The difference between non-probability and probability sampling is that non-probability sampling does not involve *random* selection and probability sampling does. Does that mean that non-probability samples aren't representative of the population? Not necessarily. But it does mean that non-probability samples cannot depend upon the rationale of probability theory. At least with a probabilistic sample, we know the odds or probability that we have represented the population well. We are able to estimate confidence intervals for the statistic. With non-probability samples, we may or may not represent the population well, and it will often be hard for us to know how well we've done so. In general,

researchers prefer probabilistic or random sampling methods over non-probabilistic ones, and consider them to be more accurate and rigorous. However, in applied social research there may be circumstances where it is not feasible, practical or theoretically sensible to do random sampling. Here, we consider a wide range of non-probabilistic alternatives.
We can divide non-probability sampling methods into two broad types: *accidental* or *purposive*. Most sampling methods are purposive in nature because we usually approach the sampling problem with a specific plan in mind. The most important distinctions among these types of sampling methods are the ones between the different types of purposive sampling approaches.

Accidental, Haphazard or Convenience Sampling

One of the most common methods of sampling goes under the various titles listed here. I would include in this category the traditional "man on the street" (of course, now it's probably the "person on the street") interviews conducted frequently by television news programs to get a quick (although non-representative) reading of public opinion. I would also argue that the typical use of college students in much psychological research is primarily a matter of convenience.
(You don't really believe that psychologists use college students because they believe they're representative of the population at large, do you?). In clinical practice, we might use clients who are available to us as our sample.
In many research contexts, we sample simply by asking for volunteers. Clearly, the problem with all of these types of samples is that we have no evidence that they are representative of the populations we're interested in generalizing to -- and in many cases we would clearly suspect that they are not.

Purposive Sampling

In purposive sampling, we sample with a *purpose* in mind. We usually would have one or more specific predefined groups we are seeking. For instance, have you ever run into people in a mall or on the street who are carrying a clipboard and who are stopping various people and asking if they could interview them? Most likely they are conducting a purposive sample (and most likely they are engaged in market research). They might be looking for Caucasian females between 30-40 years old. They size up the people passing by and anyone who looks to be in that category they stop to ask if they will participate. One of the first things they're likely to do is verify that the respondent does in fact meet the criteria for being in the sample. Purposive sampling can be very useful for situations where you need to reach a targeted sample quickly and where sampling for proportionality is not the primary concern. With a purposive sample, you are likely to get the opinions of your target population, but you are also likely to overweight subgroups in your population that are more readily accessible.
All of the methods that follow can be considered subcategories of purposive sampling methods. We might sample for specific groups or types of people as in modal instance, expert, or quota sampling. We might sample for diversity as in heterogeneity sampling. Or, we might capitalize on informal social networks to identify specific respondents who are hard to locate otherwise, as in snowball sampling. In all of these methods we know what we want -- we are sampling with a purpose.

- **Modal Instance Sampling**

In statistics, the *mode* is the most frequently occurring value in a distribution. In sampling, when we do a modal instance sample, we are sampling the most frequent case, or the "typical" case. In a lot of informal public opinion polls, for instance, they interview a "typical" voter. There are a number of problems with this sampling approach.
First, how do we know what the "typical" or "modal" case is? We could say that the modal voter is a person who is of average age, educational level, and income in the population. But, it's not clear that using the averages of these is the fairest (consider the skewed distribution of income, for instance). And, how do you know that those three variables -- age, education, income -- are the only or even the most relevant for classifying the typical voter? What if religion or ethnicity is an important discriminator? Clearly, modal instance sampling is only sensible for informal sampling contexts.

- **Expert Sampling**

Expert sampling involves the assembling of a sample of persons with known or demonstrable experience and expertise in some area. Often, we convene such a sample under the auspices of a "panel of experts." There are actually two reasons you might do expert sampling In this case, expert sampling is essentially just a specific sub case of purposive sampling. But the other reason you might use expert sampling is to provide evidence for the validity of another sampling approach you've chosen. For instance, let's say you do modal instance sampling and are concerned that the criteria you used for defining the modal instance are subject to criticism. You might convene an expert panel consisting of persons with acknowledged experience and insight into that field or topic and ask them to examine your modal definitions and comment on their appropriateness and validity. The advantage of doing this is that you aren't out on your own trying to defend your decisions -- you have some acknowledged experts to back you. The disadvantage is that even the experts can be, and often are, wrong.

- **Quota Sampling**

In quota sampling, you select people non randomly according to some fixed quota. There are two types of quota sampling: *proportional* and *non-proportional*. In **proportional quota sampling** you want to represent the major characteristics of the population by sampling a proportional amount of each. For instance, if you know the population has 40% women and 60% men, and that you want a total sample size of 100, you will continue sampling until you get those percentages and then you will stop.
So, if you've already got the 40 women for your sample, but not the sixty men, you will continue to sample men but even if legitimate women respondents come along, you will not sample them because you have already "met your quota." The problem here (as in much purposive sampling) is that you have to decide the specific characteristics on which you will base the quota. Will it be by gender, age, education race, religion, etc.?

Non proportional quota sampling is a bit less restrictive. In this method, you specify the minimum number of sampled units you want in each category. Here, you're not concerned with having numbers that match the proportions in the population. Instead, you simply want to have enough to assure that you will be able to talk about even small groups in the population. This method is the non-probabilistic analogue of stratified random sampling in that it is typically used to assure that smaller groups are adequately represented in your sample.

- **Heterogeneity Sampling**

We sample for heterogeneity when we want to include all opinions or views, and we aren't concerned about representing these views proportionately. Another term for this is sampling for *diversity*. In many brainstorming or nominal group processes (including concept mapping), we would use some form of heterogeneity sampling because our primary interest is in getting broad spectrum of ideas, not identifying the "average" or "modal instance" ones. In effect, what we would like to be sampling is not people, but ideas.

We imagine that there is a universe of all possible ideas relevant to some topic and that we want to sample this population, not the population of people who have the ideas. Clearly, in order to get all of the ideas, and especially the "outlier" or unusual ones, we have to include a broad and diverse range of participants. Heterogeneity sampling is, in this sense, almost the opposite of modal instance sampling.

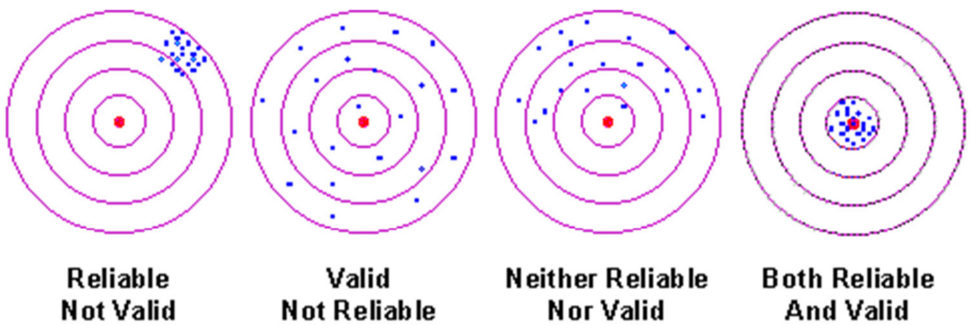
- **Snowball Sampling**

In snowball sampling, you begin by identifying someone who meets the criteria for inclusion in your study. You then ask them to recommend others who they may know who also meet the criteria. Although this method would hardly lead to representative samples, there are times when it may be the best method available. Snowball sampling is especially useful when you are trying to reach populations that are inaccessible or hard to find. For instance, if you are studying the homeless, you are not likely to be able to find good lists of homeless people within a specific geographical area. However, if you go to that area and identify one or two, you may find that they know very well who the other homeless people in their vicinity are and how you can find them.

Reliability& Validity

We often think of reliability and validity as separate ideas but, in fact, they're related to each other. Here, I want to show you two ways you can think about their relationship.

One of my favorite metaphors for the relationship between reliability is that of the target. Think of the center of the target as the concept that you are trying to measure. Imagine that for each person you are measuring, you are taking a shot at the target. If you measure the concept perfectly for a person, you are hitting the center of the target. If you don't, you are missing the center. The more you are off for that person, the further you are from the center.



The figure above shows four possible situations. In the first one, you are hitting the target consistently, but you are missing the center of the target. That is, you are consistently and systematically measuring the wrong value for all respondents. This measure is reliable, but no valid (that is, it's consistent but wrong). The second shows hits that are randomly spread across the target. You seldom hit the center of the target but, on average, you are getting the right answer for the group (but not very well for individuals). In this case, you get a valid group estimate, but you are inconsistent. Here, you can clearly see that reliability is directly related to the variability of your measure. The third scenario shows a case where your hits are spread across the target and you are consistently missing the center. Your measure in this case is neither reliable nor valid. Finally, we see the "Robin Hood" scenario -- you consistently hit the center of the target. Your measure is both reliable and valid (I bet you never thought of Robin Hood in those terms before).

What is **reliability**? We hear the term used a lot in research contexts, but what does it really mean? If you think about how we use the word "reliable" in everyday language, you might get a hint. For instance, we often speak about a machine as reliable: "I have a reliable car." Or, news people talk about a "usually reliable source". In both cases, the word reliable usually means "dependable" or "trustworthy." In research, the term "reliable" also means dependable in a general sense, but that's not a precise enough definition. What does it mean to have a dependable measure or observation in a research context? The reason "dependable" is not a good enough description is that it can be confused too easily with the idea of a valid measure (see Measurement Validity).

Certainly, when we speak of a dependable measure, we mean one that is both reliable and valid. So we have to be a little more precise when we try to define reliability.

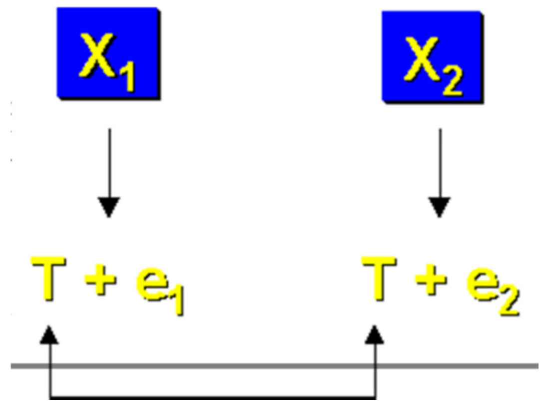
Reliability refers to the degree to which a measuring procedure gives consistent results. That is, a reliable test is a test which would provide a consistent set of scores for a group of individuals if it was administered independently on several occasions.

Reliability is a necessary but not sufficient condition for validity. A test which provides totally inconsistent results cannot possibly provide accurate information about the behavior being measured. Thus low reliability can be expected to restrict the degree of validity that is obtained, but high reliability provides no guarantee that a satisfactory degree of validity will be present.

Note that reliability refers to the nature of the test scores and not to the test itself. Any particular test may have a number of different reliabilities, depending on the group involved and the situation in which it is used. The assessment of reliability is measured by the 'Reliability Coefficient' (for groups of individuals) or the "Standard Error of Measurement" (for individuals).

In research, the term reliability means "repeatability" or "consistency". A measure is considered reliable if it would give us the same result over and over again (assuming that what we are measuring isn't changing!).

Let's explore in more detail what it means to say that a measure is "repeatable" or "consistent". We'll begin by defining a measure that we'll arbitrarily label **X**. It might be a person's score on a math achievement test or a measure of severity of illness. It is the value (numerical or otherwise) that we observe in our study. Now, to see how repeatable or consistent an observation is, we can measure it twice. We'll use subscripts to indicate the first and second observation of the same measure. If we assume that what we're measuring doesn't change between the time of our first and second observation, we can begin to understand how we get at reliability. While we observe a score for what we're measuring, we often think of that score as consisting of parts.



It's important to keep in mind that we observe the **X** score -- we never actually see the true (**T**) or error (**e**) scores. For instance, a student may get a score of **85** on a math achievement test. That's the score we observe, an **X** of **85**. But the reality might be that the student is actually better at math than that score indicates. Let's say the student's true math ability is **89** (i.e., **T=89**). That means that the error for that student is **-4**. What does this mean? Well, while the student's true math ability may be **89**, he/she may have had a bad day, may not have had breakfast, may have had an argument, or may have been distracted while taking the test. Factors like these can contribute to errors in measurement that make the student's observed ability appear lower than their true or actual ability.

OK, back to reliability. If our measure, **X**, is reliable, we should find that if we measure or observe it twice on the same persons that the scores are pretty much the same. But why would they be the same? If you look at the figure you should see that the only thing that the two observations have in common is their true scores, **T**. How do you know that? Because the error scores ($\mathbf{e_1}$ and $\mathbf{e_2}$) have different subscripts indicating that they are different values. But the true score symbol **T** is the same for both observations. What does this mean? That the two observed scores, $\mathbf{X_1}$ and $\mathbf{X_2}$ are related only to the degree that the observations share true score. You should remember that the error score is assumed to be random. Sometimes errors will lead you to perform better on a test than your true ability (e.g., you had a good day guessing!) while other times it will lead you to score worse. But the true score -- your true ability on that measure -- would be the same on both observations (assuming, of course, that your true ability didn't change between the two measurement occasions).

With this in mind, we can now define reliability more precisely. Reliability is a **ratio** or fraction. In layperson terms we might define this ratio as:

True level on the measure

The entire measure

You might think of reliability as the proportion of "truth" in your measure. Now, we don't speak of the reliability of a measure for an individual -- reliability is a characteristic of a measure that's taken across individuals. So, to get closer to a more formal definition, let's restate the definition above in terms of a set of observations. The easiest way to do this is to speak of the variance of the scores. Remember that the variance is a measure of the spread or distribution of a *set* of scores. So, we can now state the definition as:

The variance of the true score

The variance of the measure

We might put this into slightly more technical terms by using the abbreviated name for the variance and our variable names:

var(T)

var(X)

We're getting to the critical part now. If you look at the equation above, you should recognize that we can easily determine or calculate the bottom part of the reliability ratio -- it's just the variance of the set of scores we observed (You remember how to calculate the variance, don't you? It's just the sum of the squared deviations of the scores from their mean, divided by the number of scores). But how do we calculate the variance of the true scores. We can't see the true scores (we only see X)! Only God knows the true score for a specific observation.

And, if we can't calculate the variance of the true scores, we can't compute our ratio, which means ***we can't compute reliability***! Everybody got that? The bottom line is...

We can't compute reliability because we can't calculate the variance of the true scores

Great! So where does that leave us? If we can't compute reliability, perhaps the best we can do is to *estimate* it. Maybe we can get an estimate of the variability of the true scores. How do we do that? Remember our two observations, X_1 and X_2? We assume (using true score theory) that these two observations would be related to each other to the degree that they share true scores. So, let's calculate the correlation between X_1 and X_2. Here's a simple formula for the correlation:
covariance(X_1, X_2)

Sd (X_1) * sd (X_2)

Where the 'sd' stands for the standard deviation (which is the square root of the variance); If we look carefully at this equation, we can see that the covariance, which simply measures the "shared" variance between measures, must be an indicator of the variability of the true scores because the true scores in X_1 and X_2 are the only thing the two observations share! So, the top part is essentially an estimate of **var(T)** in this context. And, since the bottom part of the equation multiplies the standard deviation of one observation with the standard deviation of the same measure at another time, we would expect that these two values would be the same (it is the same measure we're taking) and that this is essentially the same thing as squaring the standard deviation for either observation. But, the square of the standard deviation is the same thing as the variance of the measure. So, the bottom part of the equation becomes the variance of the measure (or **var(X)**). If you read this paragraph carefully, you should see that the correlation between two observations of the same measure is an estimate of reliability.

It's time to reach some conclusions. We know from this discussion that we cannot calculate reliability because we cannot measure the true score component of an observation. But we also know that we can *estimate* the true score component as the covariance between two observations of the same measure.

With that in mind, we can estimate the reliability as the correlation between two observations of the same measure. It turns out that there are several ways we can estimate this reliability correlation. These are discussed in Types of Reliability.

There's only one other issue I want to address here. How big is an estimate of reliability? To figure this out, let's go back to the equation given earlier:

var(T)

var(X)
And remember that because X = T + e, we can substitute in the bottom of the ratio:
var(T)

var(T) + var(e)
With this slight change, we can easily determine the range of a reliability estimate. If a measure is perfectly reliable, there is no error in measurement -- everything we observe is true score. Therefore, for a perfectly reliable measure, the equation would reduce to:
var(T)

var(T)

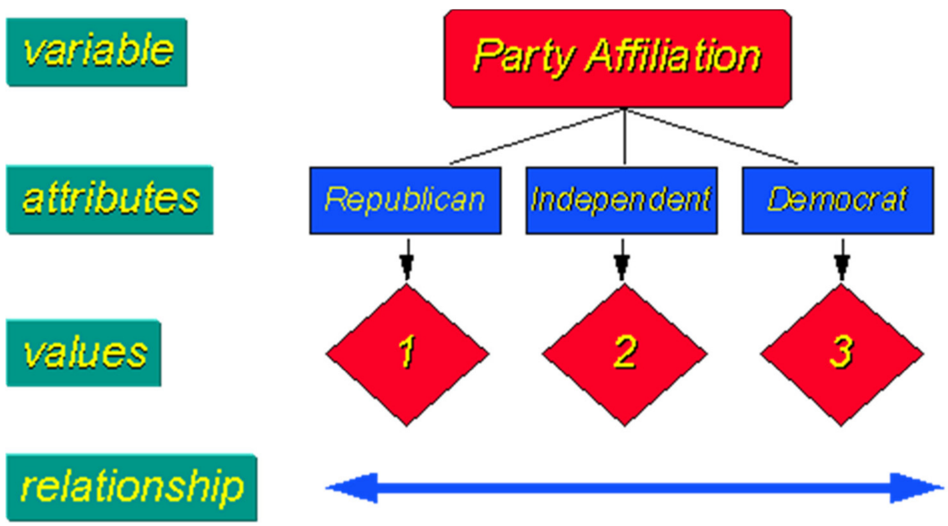
And reliability = 1. Now, if we have a perfectly unreliable measure, there is no true score -- the measure is entirely error. In this case, the equation would reduce to:

var(e)

And the reliability = 0. From this we know that reliability will always range between 0 and 1. The value of a reliability estimate tells us the proportion of variability in the measure attributable to the true score. A reliability of .5 means that about half of the variance of the observed score is attributable to truth and half is attributable to error. A reliability of .8 means the variability is about 80% true ability and 20% error. And so on.

Levels of Measurement

The level of measurement refers to the relationship among the values that are assigned to the attributes for a variable. What does that mean? Begin with the idea of the variable, in this example "party affiliation." That variable has a number of attributes. Let's assume that in this particular election context the only relevant attributes are "republican", "democrat", and "independent". For purposes of analyzing the results of this variable, we arbitrarily assign the values 1, 2 and 3 to the three attributes.



The ***level of measurement*** describes the relationship among these three values. In this case,
We simply are using the numbers as shorter placeholders for the lengthier text terms. We don't assume that higher values mean "more" of something and lower numbers signify "less". We don't assume the value of 2 means that democrats are twice something that republicans are.
We don't assume that republicans are in first place or have the highest priority just because they have the value of 1. In this case, we only use the values as a shorter name for the attribute. Here, we would describe the level of measurement as "nominal".

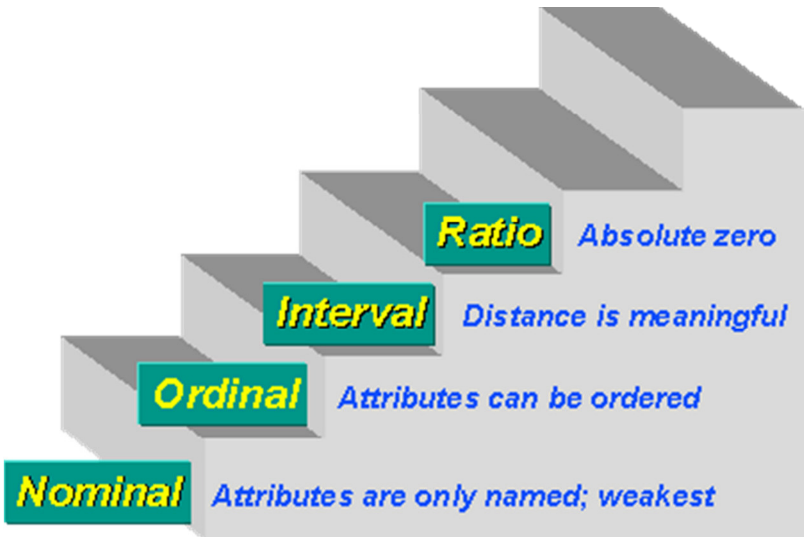
Why is Level of Measurement Important?

First, knowing the level of measurement helps you decide how to interpret the data from that variable. When you know that a measure is nominal (like the one just described), then you know that the numerical values are just short codes for the longer names. Second, knowing the level of measurement helps you decide what statistical analysis is appropriate on the values that were assigned. If a measure is nominal, then you know that you would never average the data values or do a t-test on the data.
There are typically four levels of measurement that are defined:

- Nominal
- Ordinal
- Interval
- Ratio

In **nominal** measurement the numerical values just "name" the attribute uniquely. No ordering of the cases is implied. For example, jersey numbers in basketball are measures at the nominal level. A player with number 30 is not more of anything than a player with number 15, and is certainly not twice whatever number 15 is.

In **ordinal** measurement the attributes can be rank-ordered. Here, distances between attributes do not have any meaning. For example, on a survey you might code Educational Attainment as 0=less than H.S.; 1=some H.S.; 2=H.S. degree; 3=some college; 4=college degree; 5=post college. In this measure, higher numbers mean *more* education. But is distance from 0 to 1 same as 3 to 4? The interval between values is not interpretable in an ordinal measure.



In **interval** measurement the distance between attributes *does* have meaning. For example, when we measure temperature (in Fahrenheit), the distance from 30-40 is same as distance from 70-80. The interval between values is interpretable.

Because of this, it makes sense to compute an average of an interval variable, where it doesn't make sense to do so for ordinal scales. But note that in interval measurement ratios don't make any sense - 80 degrees is not twice as hot as 40 degrees (although the attribute value is twice as large).

Finally, in **ratio** measurement there is always an absolute zero that is meaningful. This means that you can construct a meaningful fraction (or ratio) with a ratio variable. Weight is a ratio variable. In applied social research most "count" variables are ratio, for example, the number of clients in past six months. Why? Because you can have zero clients and because it is meaningful to say that "...we had twice as many clients in the past six months as we did in the previous six months."

It's important to recognize that there is a hierarchy implied in the level of measurement idea. At lower levels of measurement, assumptions tend to be less restrictive and data analyses tend to be less sensitive. At each level up the hierarchy, the current level includes all of the qualities of the one below it and adds something new. In general, it is desirable to have a higher level of measurement (e.g., interval or ratio) rather than a lower one (nominal or ordinal).

Scaling

Scaling is the branch of measurement that involves the construction of an instrument that associates qualitative constructs with quantitative metric units.

Scaling evolved out of efforts in psychology and education to measure "un-measurable" constructs like authoritarianism and self-esteem. In many ways, scaling remains one of the most arcane and misunderstood aspects of social research measurement. And, it attempts to do one of the most difficult of research tasks -- measure abstract concepts.

Most people don't even understand what scaling is. The basic idea of scaling is described in General Issues in Scaling, including the important distinction between a scale and a response format. Scales are generally divided into two broad categories: uni-dimensional and multidimensional. The one-dimensional scaling methods were developed in the first half of the twentieth century and are generally named after their inventor. We'll look at three types of uni-dimensional scaling methods here:

- Thurstone or Equal-Appearing Interval Scaling
- Likert or "Summative" Scaling
- Guttman or "Cumulative" Scaling

In the late 1950s and early 1960s, measurement theorists developed more advanced techniques for creating multidimensional scales. Although these techniques are not considered here, you may want to look at the method of concept mapping that relies on that approach to see the power of these multivariate methods.

REFERENCES

March 3, 2011 from http://www.jllonline.co.uk/journal/jllearn/vol1no2_home.html
Roth W. M (2002) Editorial power/authorial suffering Research, in Science Education, 32 (2), 215:240,
Sullivan, S (1996). Scholarly Publishing: Trash or Treasure. Australian Academic &
Research Libraries, 27 (1): 40-46.
Syed, Z (2003). TESOL in the Gulf: The socio-cultural context of English language teaching in the Gulf. TESOL Quarterly, 37 (2): 337-341.
Turner L & Mairesse J (2005) Individual productivity differences in public research: How Important is non-individual
Determinants? An econometric study of French physicists publications and citations. Retrieved March 31, 2011 from http://piketty.pse.ens.fr/fichiers/semina/lunch/Turner2005.pdf.
UNESCO (2005) English for all – The quality imperative.
World Bank (2008) the road not travelled: Education reform in the Middle East and Africa.
Washington DC. Retrieved March 15, 2009 from http://www.worldbank.org/html/extpb/educat.htm.
Zughoul M. (2003) Globalization and EFL/ESL. Pedagogy in the Arab World,
Pasigna, A.I. (2001) *Continuous Assessment Remediation and Enrichment*. Washington, DC: American Institute for Research.
Rust.G.H, (2002)" The Impact of Assessment on Student Learning: How can the Research Literature Practically Help to Inform the Development of Departmental Assessment Strategies and Learner Center Assessment Practices?"
Retrieved from http://www.a/h.sagepub.com/cgi/content/refs/3/2/145.
Shamim, F. *ET.al.2007*. *Maximizing Learning in Large Classes: Issues and Options*. Addis Ababa:Master Printing Press Plc.