Service-Learning Research Primer

© 2010 Learn and Serve America's National Service-Learning Clearinghouse

Authors: Kathryn S. Steinberg, Robert G. Bringle, and Matthew J. Williams

Printed in the United States of America

Suggested citation for this document:

Steinberg, K. S., Bringle, R. G., & Williams, M. J. (2010). *Service-learning research primer*. Scotts Valley, CA: National Service-Learning Clearinghouse. http://servicelearning.gov/filemanager/download/Service-Learning_Research_Primer.pdf

Contents

Chapter 1
The Nature of Scientific Research

The use of service-learning as a pedagogy in higher education classes has blossomed over the past 20 years in both undergraduate and graduate courses (Campus Compact, 2006). There is an acute need for high-quality research on service-learning outcomes across institutions, faculty, students, and communities (Bringle & Hatcher, 2000). For, as Eyler noted,

> [Service-learning research] is neither precise nor robust enough to guide decision making about practice. Our success at implementation has outstripped our knowledge of what works best. For a field that engenders as much passion in practitioners and that we believe transforms students by engaging their hearts as well as their minds, there is remarkably little evidence of strong impact and even less evidence about the kinds of practices that lead to the effects we desire. (Eyler, 2002, p. 5)

Although research on service-learning represents a nascent field of endeavor, a number of organizations and resources have been developed to assist interested persons in their research activities. For example, Campus Compact, Learn and Serve America's National Service-Learning Clearinghouse (NSLC), Community-Campus Partnerships for Health, and the International Association for Research on Service-Learning and Civic Engagement all have developed resources that are available on their websites to assist researchers and program planners and provide opportunities for disseminating research results. RMC Research has made available online the Compendium of Assessment and Research Tools (cart.rmcdenver.com/). The International Association for Research on Service-Learning and Community Engagement (IARSLCE) was launched "to promote the development and dissemination of research on service-learning and community engagement internationally and across all levels of the education system" (International Association for Research on Service-Learning and Community Engagement, 2009). IARSLCE publishes a series of volumes, *Advances in Service-Learning Research*, developed from the annual research conference. The *Michigan Journal of Community Service Learning* has been a premier resource for disseminating research since 1994. A number of other institutions, including government agencies (e.g., the Corporation for National and Community Service), research centers at institutions of higher education (e.g., CIRCLE at Tufts University), and funding organizations (e.g., the Spencer Foundation) have supported research on service-learning and community engagement. NSLC provides a rich set of resources to guide the development, execution, and dissemination of research.

Yet despite these resources, there is still a need for information and resources on how to conduct high quality and rigorous research on service-learning. Too often faculty, teachers, and other researchers who utilize service-learning pedagogy in their classrooms may be experts in conducting research in their own field or discipline, but are newcomers to educational research.

Thus, too frequently they are unfamiliar with the literature base in service-learning, the research methodologies that are appropriate in this field, measurement procedures, and online resources that are available. This research primer is designed to address that need.

Service-Learning

Service-learning is defined as a "course-based, credit bearing educational experience in which students (a) participate in an organized service activity that meets identified community needs, and (b) reflect on the service activity in such a way as to gain further understanding of course content, a broader appreciation of the discipline, and an enhanced sense of personal values and civic responsibility" (Bringle & Hatcher, 2009, p.38). The service component can be an option within a course, required of all students, a one-time service project, a disciplinary capstone course, or a fourth credit option in a three credit hour course (Enos & Troppe, 1996). Well-designed service-learning courses engage students in service activities that are mutually beneficial to community stakeholders (e.g., agency, recipients, community) and meet the educational objectives of the course. The educational outcomes are developed through reflection activities (e.g., journals, small group discussions, directed writing) that link the service experience to learning objectives, are guided, occur regularly, allow feedback and assessment, and include the clarification of values (Ash, Clayton, & Atkinson, 2005; Ash & Clayton, 2004; Bringle & Hatcher, 1999; Eyler, Giles, & Schmiede, 1996; Hatcher & Bringle, 1997). In addition, students enrolled in service-learning classes do not receive academic credit for engaging in community service; rather, they receive academic credit for the learning that occurs as a result of the service experience (Howard, 1993).

In addition, one goal of integrating service into academic studies is to develop a life-long habit of civic involvement and community service. Unlike many forms of experiential learning, which focus on pre-professional training (e.g., internships; Furco, 1996; Perold, 1998), one of the defining attributes of service-learning is that, along with academic learning, it also aspires to enhance students' civic growth (Annette, 2003; Ash et al., 2005; Battistoni, 2002). Thus, in addition to "serving to learn," service-learning intentionally focuses on "learning to serve." Although developing good citizens is not a new role for higher education, and there are numerous pedagogical approaches for civic learning (e.g., classroom instruction on civics, moderated discussions of current events, student governance and community activities, simulations; Levine, 2003), the emergence of service-learning has heightened attention to the nuances of the civic domain and social responsibility as a set of intentional educational objectives to be addressed seriously in higher education (Astin & Sax, 1998). Even though, as Dionne and Drogosz (2003) note, "citizenship cannot be reduced to service" (p. 25), service-learning needs to be better understood as a means for teaching toward civic learning objectives.

The Nature of Inquiry

There are numerous forms of inquiry and evidence that people rely on to make inferences, establish practice, modify behavior, and persuade others, including the following:

1. **Intuition** can be used to reach a conclusion. One of the limitations of intuition is that it is based on a private process that does not allow others to evaluate the evidence or the process upon which the inference was made.

2. **Experience** provides important information. Persons trust their senses and their experiences. "If I have (not) experienced something, then it must (not) be true!" Experiential knowledge has limitations that include the sample of experiences to which people have been exposed, and the limitations of the senses including processing sensory information, and accurately recalling the information.

3. **Observation** contains information that includes the experiences, inferences, behaviors, and verbal representations of others as well as self. Casual observations contain limitations imposed by unrepresentative samples, biases of perception and attention, imprecise conceptual frameworks, and difficulty in making clear causal statements about why a behavior occurred.

4. **Dogma, authority, and opinion of experts** capture a basis of inferences that refer to, for example, religious, political, and social knowledge bases that determine and influence belief systems and inferences (e.g., "It's true because the Bible [or the President or an expert] says so").

5. **Consensus** bases inferences on what others generally think or are believed to think. A limitation of consensus (as with dogma, authority, and opinion of experts) is that consensus groups may have biases in perception and frameworks of reasoning.

6. **Logic** presumes that if one can reason correctly (accurately), then valid conclusions will be reached. The validity of reasoning by logic is limited by the nature of the premise. Although sound logic may contribute to good research, it is not the same as research.

There are different formal means through which information can be accumulated and analyzed. For example, philanthropy can be analyzed from the following paradigmatic perspectives: philosophical, historical, economic, literary, dramaturgical, educational, linguistic, or psychological. Each of these paradigms uses its own methods, language, conceptual frameworks, and tools. Each also contributes to the overall understanding of a phenomenon.

Thus, research using paradigms of traditional science are not better than other modes of inquiry; rather they are better viewed as complementing other methods of inquiry.

The scientific method itself is a collection of different paradigms. For example, eating would be approached with very different methods, tools, and conceptual perspectives by an anthropologist, sociologist, physiological psychologist, social psychologist, biochemist, and neurologist. Colloquially, scientific research is assumed to describe the process of collecting data. However, people have always collected information (data) through observation and experience, and they have used consensus, dogma, logic, and intuition to make inferences about highly significant and relevant issues (e.g., disease, earthquakes, kindness). Research is distinguished from casual observation and other forms of inquiry by how its methods provide information that can meaningfully contribute to an understanding of phenomena by providing a basis for deductions and generalizations. Research aspires to be empirical (i.e., rely on observable events), to conduct information gathering according to a set of procedural rules (i.e., to be systematic), to control for alternative explanations, to be public and open to scrutiny (i.e., the methods are available for critical evaluation), and to be amoral or value free (i.e., acknowledge and attempt to overcome biases, be objective). Thus, research is the systematic collection and synthesis of empirical information that supports inferences or conclusions.

Quantitative approaches to scientific research emphasize testing hypotheses deduced from broad, abstract theories. Hypotheses traditionally are used to test the robustness and adequacy of theories across diverse sets of examples. Qualitative approaches to research emphasize the interpretation of in-depth meaning (e.g., how, why) that is available from one or multiple sources that may be purposively selected and studied.

In addition to its methods, the effectiveness and meaningfulness of scientific research is tied to its theoretical context (See Figure 1). Thus, it is a myth that the purpose of research is to collect data merely for its own sake. Research is better viewed as the collection of data in the service of theory.

The Nature of Theory

Theory is a set of interrelated propositions about constructs. Theories are conceptual frameworks that aid in organizing and predicting phenomena. Again, just as everyone collects data, both lay persons and researchers have theories. According to McGuire (1980), formal theories "are distinguished from ordinary thought by the use of more explicit definitions and more abstract and more formally interrelated principles" (p. 53). Theories vary in their specificity and their units of analysis. Deutsch and Krauss illustrate this through the metaphor of a net:

> Theory is a net man weaves to catch the world of observation—to explain, predict, and influence it. The theorists . . . have woven nets of different sorts. Some are all-purpose nets meant to catch many species of behavior, others are

clearly limited to a few species; some have been constructed for use near the surface, others work best in the depths. (Deutsch & Krauss, 1965, p. vii)

Their analogy illustrates that the utility, meaningfulness, and validity of a theory will depend upon its application and frame of reference. Shaw and Costanzo (1982) identify three characteristics that they consider necessary in the evaluation of the quality of a theory:

1. Logical and internal consistency;
2. Agreement with known data;
3. Testability.

In addition, they identify three characteristics that are desirable:

1. Simple in its presentation;
2. Economical in its ability to explain phenomena;
3. Consistent with related theories.

Theories represent templates through which phenomena are interpreted using language and thought. But they are not "merely theories" in the sense that they are inconsequential. McGuire poignantly states how indispensable theories are:

What makes theorizing a tragedy is not that our theories are poor but that, poor as they are, they are essential, for we cannot do without them. The ubiquity of formal and informal theorizing demonstrates its indispensability. To cope with reality we must reduce it to the oversimplified level of complexity that our minds can manage and distort it into the type of representations that we can grasp. We are reduced to groping for theories that are happy instances of brilliant oversimplification whose elected ignorances and distortions happen to be incidental to the matter under consideration, so that within the momentary situation the theory's apt focusing of our creative and critical appraisal yields gains that outweigh the losses caused by its oversights and distortions. (McGuire, 1980, p. 54)

The value of theories is that they clarify and simply information, inferences, and decisions by providing a context within which questions can be asked and answered. Theories can provide a rich set of heuristics through which to explore auxiliary phenomena, boundary conditions, and alternative conceptual frameworks. In doing so, theories suggest additional means for analyzing phenomena, a context for subsequent predictions, and a basis of relating research findings to other research and theories. Bringle (2003) has contended that research on

service-learning has suffered from a lack of attention to theory and has suggested ways in which theories can be borrowed from cognate areas or developed to improve service-learning research.

Interplay Between Theory and Research

Theory and research are equally important to the process of accumulating knowledge through the scientific method (Bringle, 2003; Bringle & Hatcher, 2000). The process can begin at different points on the diagram (Figure 1). It may start with a preliminary theory that, through the deductive process, generates testable hypotheses that are evaluated through research, the results of which produce decisions about the theory (e.g., supported, refuted, need to revise). Alternatively, specific observations may be used to generalize principles that are conceptually developed into a theory that then guides subsequent research that evaluates research questions and deduced hypotheses. The presumption is that, in every case, there is a symbiotic relationship between theory and research, such that theory guides the research process, and research results arbitrate an evaluation of the appropriateness of the theory (e.g., supported, needs modification, refuted). This is true whether the research is quantitative or qualitative in nature. Figure 1 illustrates the importance of two types of connections between research and theory, namely that the relationship involves a cycle of both inductive and deductive processes.

Figure 1. The Relationship between Research and Theory

Theory

Induction

Deduction

Research

- **Quantitative**
- **Qualitative**
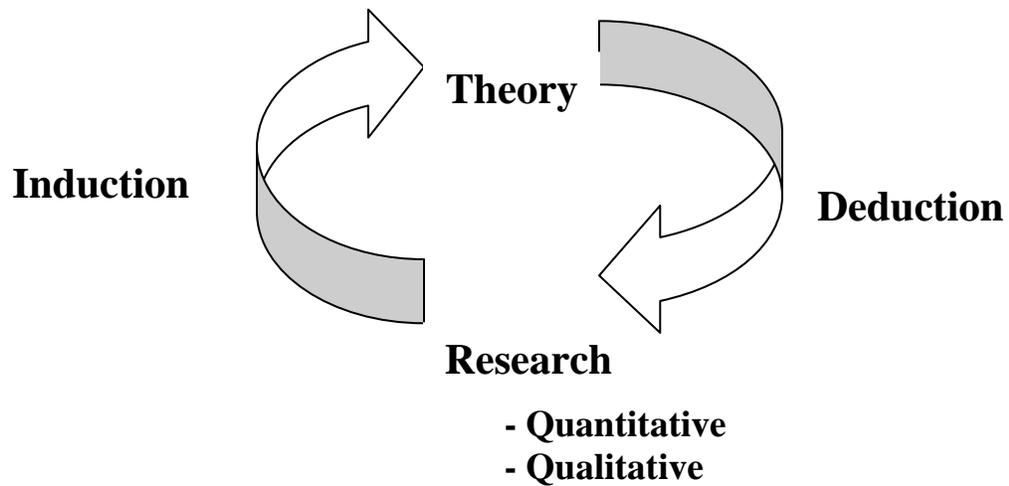
Thus, theories are comprised of statements about the nature of constructs, their manifestations, and the relationships between constructs. **Constructs** are abstract or hypothetical entities. Critical thinking is a construct. No one can see critical thinking. Theorists can map the conceptual domain and identify attributes that are presumed to be indicative of good or poor critical thinking, but the construct itself does not exist in a tangible way. The manifestations of the construct (e.g., verbal or behavioral manifestations of critical thinking) may be apparent, help

differentiate among individuals, and allow one to rank persons on some attribute associated with critical thinking, but the construct itself is not directly accessible. ***Variables***, or the phenomena of interest in a research study, are the concrete manifestations of constructs that are either (a) ***quantitative***, in that they vary in intensity or degree, or (b) ***qualitative***, in that they differ in kind.

Operationalization refers to a statement about the specific way in which a quantitative or qualitative variable is measured, observed, documented, or manipulated in research. The progression from construct to variable to operationalization, thus, is a deductive process that goes from more abstract to more concrete. For example, reflection is assumed to be an integral component of designing a successful service-learning class. Reflection is defined as the "intentional consideration of an experience in light of particular learning objectives" (Hatcher & Bringle, 1997, p. 53). As such, reflection is a construct. There are many ways in which reflection can occur, including journals, directed writings, critical incident papers, group discussions, and portfolios. There are also dimensions on which these methods can vary (e.g., structured vs. unstructured). The implementation of these forms of reflection, their operationalization, could be a quantitative variable (e.g., some students are asked to write ten pages of journal entry whereas other students are asked to write 100 pages). Or, the operationalization of reflection could be a qualitative variable in that the activities differ in kind (e.g., some students are asked to write 20 pages of journal entry whereas other students engage in a series of group discussions). Thus, not only are there multiple variables associated with a construct, but there are also many ways to operationalize any one of the variables. In quantitative research, the key is to operationalize the construct in such a way as to be able to evaluate a ***hypothesis***, which is a tentative statement about the expected result. Classically, qualitative research is characterized by a discovery-type approach in which no prior constraints are made on the observation methods or study results. In practice, most qualitative researchers do at least make an initial outline of what type of instruments and procedures they will use, and what types of questions they are seeking to answer.

In order to evaluate a hypothesis using the quantitative approach, the researcher must structure the data collection in such a way that inferences can be made. This requires adequate design and implementation of the research procedures (***internal validity***), utilization of reliable and valid measurements or observations, conducting appropriate analysis of the data (***statistical validity***), and making appropriate inductive inferences about the pattern of results to be able to draw conclusions about its practical and theoretical implications (***external validity***).

The remainder of this primer will outline the processes and procedures that are useful in conducting research on service-learning. Chapter 2 describes the process of designing service-learning research, including the research cycle, qualities of good research, and differentiating between research and evaluation. The chapter will also provide a description of how to define research variables and give a list of sample variables that might be investigated in service-learning research. Next, the chapter focuses on research designs, common problems in service-learning research and how to address them, and ethical issues.

Chapter 3 focuses on measurement issues in service-learning research. First, a description of the most common types of assessment tools used in service-learning research, such as surveys, focus groups, and content analysis, is provided. Following this is a discussion of the characteristics of good measurement instruments (reliability, validity, and practical concerns). Then the chapter presents the pros and cons of an important decision point, whether to use existing tools, to adapt from those that have already been developed, or to create new tools to fit a specific purpose. Chapter 3 concludes with a list of resources for conducting online surveys.

Data analysis and interpretation is the focus of Chapter 4. The chapter starts with an introduction that includes a list of common pitfalls in analyzing and interpreting data, and moves to the most commonly used forms of quantitative analysis, descriptive and inferential statistics, are briefly discussed along with a special focus on the procedures more frequently used in service-learning research. This section is intended to be an introduction to the topic; for more in-depth and specific information the reader is urged to consult a statistics manual and/or another researcher who has experience or expertise in this area. Following this is a discussion of generalizability and the inductive process of research interpretation, drawing conclusions, and making recommendations.

Chapter 5 begins with a list of potential avenues for disseminating research, and then we discuss qualities of good quantitative research articles. Following this is an annotated list of publication outlets for research on service-learning. Finally, the appendices provide an annotated list of service-learning research resources on the internet, online resources on research methodology and statistics, and a listing of potential funders for service-learning research projects.

Chapter 2
Designing Service-Learning Research

Research versus Evaluation: Scope of Conclusions

> Research, especially fundamental or basic research, differs from evaluation in that its primary purpose is to generate or test theory and contribute to knowledge for the sake of knowledge. Such knowledge, and the theories that undergird knowledge, may subsequently inform action and evaluation, but action is not the primary purpose of fundamental research. (Patton, 2002, pp. 10-11)

There are several types of data collection that are involved when studying service-learning education and these vary in the nature and scope of the conclusions that are drawn. The focus of this *Research Primer* is on **research** on service-learning. Other forms of assessment focus on classroom assessment, self-assessment, and course or program evaluation. These are summarized briefly here and illustrated in Table 1.

Classroom assessment (grading) and self-assessment techniques include student quizzes and tests, homework, reflection activities, and faculty teaching portfolios. These measurement tools allow for conclusions to be drawn about the learning experience of a specific student or faculty member. The purpose of course or program evaluation is to gather information on the process (implementation) and outcomes of a specific course or program. Conclusions are generally limited to that specific program or course. Data collected during **research**, on the other hand, lead to conclusions that can be generalized to other similar courses, programs, and/or service-learning in general.

The Research Cycle

As described in Chapter 1, the cycle of research encompasses theory. "All research, both quantitative and qualitative, is most beneficial when the design of research is guided by a theory and when the information that is gained through data collection is relevant to supporting, developing, refining, and revising a theory" (Bringle & Hatcher, 2000, p. 69). The research cycle is illustrated in Figure 2. The basic process is the same whether the researcher uses a quantitative or qualitative approach. Qualitative research, however, involves a more iterative process than quantitative research, since the investigator uses an interpretative, discovery-type approach in determining procedures to be used and data to be gathered. That is, the qualitative researcher may start out looking at one type of document, noticing particular patterns and themes, then decide to switch to oral interviews to obtain a different perspective on the themes. This iterative process (represented by the arrows in Figure 2) might be repeated several times before the researcher creates a synthesis and interprets the data.

Table 1. Purposes of Assessment in Service-Learning Education

(Adapted from Bringle & Hatcher, 2000)

Purpose	Focus of Inquiry	Examples of Questions or Information Gathered	Scope of Conclusions
Classroom Grading, Self-assessment	Individual student, faculty, or community member	How did this student perform in my class? What did this student learn? What did I learn from teaching this course?	Specific to the individual student, course, and faculty member
Process Evaluation	Specific course, class, or program	How well was the program implemented? What worked well in this class? How could it be improved?	Specific to that class, course, or program
Outcomes Evaluation	Specific course, class, or program	What could the students do at the end of the course? What was the level of moral development at the end of the program? What were the reading skills of the tutored children?	Specific to that class, course, or program
Qualitative Research	May start with specific case, but then extends across cases, groups of students, multiple classes, courses, or programs	How did faculty integrate civic learning objectives into course syllabi?	Relevant to other (similar) courses, programs, or possibly service-learning in general
Quantitative Research: Correlational	Across groups of students, or multiple classes, courses, or programs	What is the relationship between prior service experience and students' efficacy at the end of a course?	Relevant to other (similar) courses, programs, or possibly service-learning in general
Quantitative Research: Experimental or Theory Testing	Across groups of students, or multiple classes, courses, or programs	Which type of reflection activities (highly structured or unstructured) lead to better student retention of course content?	Relevant to other (similar) courses, programs, or possibly service-learning in general

Figure 2: The Research Cycle

Ask question
Review theories and prior studies
Refine question

Define variables of interest (operationalize)

Design study-- methdology, measurement tools

IRB approval if needed

Gather data

Analyze and interpret data

Dissemination of results
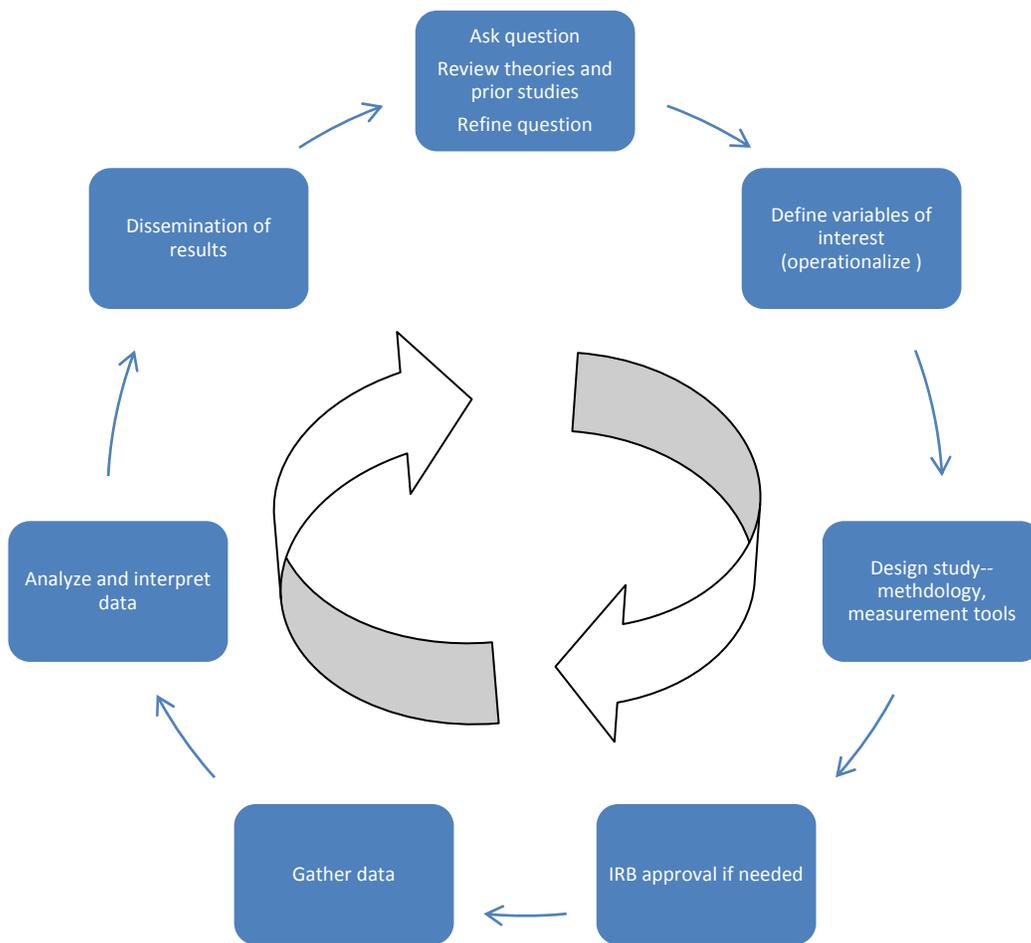
Whether qualitative or quantitative in nature, the research process itself may start out in an informal manner. For example, a faculty member may make an observation in class that leads to curiosity about whether including particular reflection exercises leads to learning a particular concept better. A review of the literature reveals different theories of active learning pedagogies, leading to refinement of the research question. The next step is defining the specific variables of interest (e.g., types of reflection) and how to manipulate and measure the outcomes (e.g., academic learning). These steps include determining the procedures to conduct the study (e.g., whether to measure learning through multiple choice exams, by conducting interviews, or through content analysis of student reflections; Will there be a control group of students?; Will learning be assessed at the beginning and end of the course?). After getting required clearance to conduct the project from the local Institutional Review Board, the researcher proceeds with data collection. The strategy used to conduct the analysis of the data will be linked with the research design and will determine the patterns of the results that are most relevant to evaluating the theoretical context of the research questions. In order for others to know about the project, the researcher needs to summarize the study in a report for dissemination. Dissemination can take a variety of forms, including an oral report to a local nonprofit agency, a presentation at a professional conference, and a peer-reviewed journal article. Frequently the research project will lead the investigator to be curious about other related questions, which starts the research cycle again.

Qualities of Good Research

Research, including experimental studies, is directed at understanding why a course, program, or experience produced a particular result. Thus, whereas program evaluation gives useful information about a particular program and its outcomes, research contributes to a knowledge base that informs others about future program design and practice by evaluating the usefulness and scope of a theory that is the basis for the program or intervention.

Furco (in Gelmon, Furco, Holland, & Bringle, 2005) notes that good quantitative research, (a) provides a theoretical frame (see Bringle, 2003), (b) uses scientific design (e.g., experimental method, such as analysis of covariance to control for pre-existing differences) to control for extraneous explanations and allow causal inferences (see Bringle & Hatcher, 2000), (c) uses measurement techniques that possess demonstrable validity and reliability (see Bringle, et al., 2004), (d) uses appropriate statistical procedures for analysis, and (e) generalizes beyond the idiosyncratic case so that others can learn from the results (see Bringle & Hatcher, 2000). Bringle (in Gelmon et al., 2005) posited that convincing research, whether quantitative or qualitative, involves:

- Guidance or grounding in theory
- Clear constructs
- Control for differences among groups (quantitative research)

- Multiple indicators
- Multiple methods
- Converging results across different methods
- Confidence in conclusions
- Implications for teaching and learning in general

Defining Research Variables (Operationalization)

An important step in designing all quantitative research projects is defining or identifying the variables that will be manipulated, measured, described, or controlled. Although qualitative researchers do not define variables to the same extent that quantitative researchers do, they still must outline what kinds of phenomena they are studying. The major types of variables, or phenomena of interest, are described briefly here, with common examples from service-learning research provided. These are presented in terms of labels from the quantitative research approach, but the qualitative tradition includes analogous examples.

- ***Independent Variable (IV):*** A variable that is *selected or controlled* by the researcher, to determine its relationship to the observed outcome of the research—also called explanatory, predictor, or manipulated variable. A common example is whether or not a course section involves service-learning pedagogy. The nature of what is varied should be carefully described so that the attributes of the different interventions or experiences are clear.
- ***Dependent Variable (DV):*** The variable being measured as an outcome—also called outcome, response, criterion, or explained variable. Many examples of dependent variables (variables of interest) are presented in Table 2.
- ***Intervening (Mediating) Variables:*** a hypothetical concept that attempts to explain the relationship between the independent and dependent variables (Baron & Kenny, 1986). Mediating variables, also called process variables, explore why the independent variable is linked to the dependent variable. For example, this might be a concept such as altruism or social responsibility that is presumed to explain why a service-learning course influenced subsequent volunteer behavior. There are statistical methods for evaluating the role of a mediating variable (see David A. Kenny's page on mediation: davidakenny.net/cm/mediate.htm).
- ***Moderator Variable:*** A variable that is related to the direction or strength of the relationship between the independent and dependent variables (Baron & Kenny, 1986). A moderator variable may be qualitative (such as student gender, type of community organization, or type of college) or quantitative (e.g., number of service visits). In addition, it may be related to the strength or the direction of a correlation, or it may interact with the independent variable and the dependent variable. In either case, a moderator variable describes an "it depends" relationship (e.g., the strength of the

correlation between two variables depends on the past volunteer experience of the student). Generally, moderator variables are variables that exist prior to data collection, as opposed to mediating variables that are assumed to occur during the phenomena being studied.

Research Designs

There are different ways to structure data collection in research. These procedures include both measurement issues and design issues. In both cases, the procedures can be sorted into quantitative and qualitative approaches. Generally speaking,

- ***Qualitative research*** focuses on analysis of documents, artifacts, words, pictures, and other non-numerical data. The approach is descriptive, interpretative, and subjective in nature.
- ***Quantitative research*** focuses on analysis of numerical data from quantitative variables. The approach often follows the scientific method of data collection by using designs that permit various levels of confidence in making causal inferences.

Although there are many adherents to each approach, some have posited (e.g., Trochim, 2006) that the dichotomy is actually false, at least as far as the *data* that are collected. For example, researchers who are by nature inclined toward the quantitative approach may utilize interviews or focus groups to explore ideas, detail theories, or develop questionnaires. Investigators who prefer a more qualitative approach may quantify interview responses into categories that are coded numerically and statistically summarized. However, the *assumptions* and philosophical approach of quantitative researchers are different from that of qualitative researchers. Many researchers (including those at the IUPUI Center for Service and Learning) agree that one approach is not inherently better than the other, and that a ***mixed-method approach*** is best, capitalizing on the strengths, and compensating for the weaknesses of each method. Because of the sometimes dramatic differences in approaches to research, however, mixing quantitative and qualitative methods without a clear rationale and purpose does not necessarily lead to better evidence to support research questions.

For more information on the strengths and weaknesses, pros and cons of the quantitative versus qualitative approaches, see the following web pages:
- Qualitative versus Quantitative Research: Key Points in a Classic Debate (wilderdom.com/research/QualitativeVersusQuantitativeResearch.html)
- The Assumptions of Qualitative Designs (www.gifted.uconn.edu/siegle/research/Qualitative/qualquan.htm)
- The Qualitative Debate (www.socialresearchmethods.net/kb/qualdeb.php)

Table 2. Sample Dependent Variables (Variables of Interest) in Service-Learning Research

Student Outcomes:

Academic:
--Learning
--Cognitive processes
--Critical thinking
--Persistence and retention
--Achievement and aspirations

Civic and Social Responsibility:
--Commitment to community
--Aspirations to volunteer
--Empathy
--Philanthropic behaviors
--Civic-minded professional

Life Skills:
--Racial tolerance
--Cultural understanding
--Self-efficacy
--Problem solving
--Communication skills
--Leadership

Personal Development:
--Moral development
--Self-concept
--Motives, attitudes, and values
--Career clarification

Faculty and Course Variables:

Teaching:
--Teaching methods
--Curriculum changes
--Grading techniques
--Barriers and facilitators

Professional Development:
--Job motivation and satisfaction
--Roles and responsibilities
--Scholarship
--Leadership

Community Variables:

Organizational:
--Type and variety of services
--Number of clients served
--Organizational capacity
--Program strategies
--Economic impact
--Networks, social impact

Community:
--Partnerships with university
--Impact on community residents
--Satisfaction with partnerships
--Sustainability of partnerships

Institutional Outcomes:
--Faculty interest and involvement in service-learning
--Relationship and involvement with external community
--Number and variety of service-learning courses offered
--Infrastructure for service-learning
--Campus mission, vision, strategic planning
--Faculty development investment
--Promotion and tenure policies
--Resource acquisition and allocation

Common Problems in Service-Learning Research

There are a number of problems that are frequently seen in research on service-learning, civic involvement, and community engagement. These are summarized in Table 3 and the discussion below.

Table 3. Common Problems in Service-Learning Research

- Small sample sizes

- Correlation ≠ Causation

- Self-selection bias (non-random assignment)

- Social desirability bias

- "Creaming the crop"

- Lack of controls or comparison groups

- Lack of generalizability (external validity)

- Not connecting to theory or knowledge base

- Lack of common definition for service-learning and other terms

- Measures are mostly self-report type

- ***Small sample sizes:*** Small sample size limits the reliability of the data, making it difficult to have confidence in the results and their implications. Examples: drawing conclusions about all service-learning students by interviewing eight seniors about one class; limiting a study to one section of one course in one semester; conducting a single case study with limited data, and then not conducting cross-case analysis to increase understanding and generalizability. Generally, effects and relationships that are found for larger samples permit more confidence in generalizing from the results to other groups.

- ***Correlation ≠ Causation***: Researchers sometimes conduct a correlational study but draw inappropriate causal (cause and effect) conclusions. For example, correlating hours of service at a site to attitudes about diversity, and then concluding, "serving more hours at a homeless shelter caused students to have more open attitudes about diversity." Without additional evidence or basis for making this causal statement, a more appropriate

inference would be something like, "students who served more hours at a homeless shelter had higher scores on the diversity scale." Correlations and causal statements are discussed further in the section on Non-Experimental Designs.

- *Self-selection bias (non-random assignment)***:** This is one of the most common problems seen in service-learning research. In most colleges and universities, (and often in secondary education) service-learning courses are not required for graduation; in addition, service-learning may be optional in a course. Thus, students *select* to be in those courses or choose those options. When conducting research, this self-selection of participants into experiences creates the problem of non-random assignment of students to a service-learning group versus a non-service-learning group and confounds the researcher's ability to determine why the students were different at the end of the experience. (See further discussion under Quasi-Experimental Designs.)

- *Social desirability bias***:** This represents a common problem in the measurement of knowledge, skills, attitudes, and behaviors related to service-learning and civic engagement. The difficulty surfaces when the behaviors and attitudes that the research wants to measure are "socially desirable" (e.g., civic-mindedness, social responsibility) and students are inclined to make themselves look good when they present responses. Researchers sometimes try to counteract this bias by including neutral or negatively-worded items in a survey or interview protocol or writing items in ways that control for the bias.

- *"Creaming the crop"***:** This problem occurs in research involving *only* students who are interested in or involved in service-learning, community service, or volunteering. The problem occurs when the investigator over-interprets or over-generalizes the results to draw conclusions about a larger group of students (e.g., all students in freshmen writing, all college students).

- *Lack of controls or comparison groups***:** Many quantitative and qualitative studies do not include adequate control or comparison groups that contrast one intervention (e.g., service-learning) with other interventions (e.g., research paper) in ways that would permit appropriate conclusions. (See more detailed discussion under Experimental and Quasi-Experimental Designs.)

- *Lack of generalizability (external validity)***:** In quantitative research, poor research design or sampling issues lead to results that cannot be generalized or applied to other situations or populations. In either qualitative or quantitative research, the nature of some studies limits the usefulness of the conclusions for other contexts. For example, research that consists of a program description may be useful for answering local questions or problems, but might not add significantly to the broader knowledge base of service-learning research. To address this problem we recommend conducting cross-case or comparative analysis to increase understanding and generalizability by searching for themes and patterns across several cases (Patton, 2002). Limitations of generalizability can apply to many aspects of the research (e.g., sampling, nature of the intervention,

context-specific elements, and measurement procedures). Generalizability is enhanced when the sample of respondents is heterogeneous (e.g., age, type or discipline of the service-learning course, type of institution). Limiting the generalization to reflect the restrictions of the sample and the study increases confidence in the research conclusions (Bringle, Phillips & Hudson, 2004).

- *Not connecting to theory or knowledge base*: Research on service-learning too seldom is cumulative across studies in meaningful ways. Rather, the field has been accumulating isolated evaluations of specific courses that have limited implications to other courses and broader practice. More research needs to have interventions and outcomes linked in systematic ways to theory. When this is done, there will be a basis of comparing and contrasting results and better understanding why outcomes were obtained or not obtained.

- *Lack of common definition of terms*: One difficulty in comparing results is that there may be no common agreement on definitions (e.g., service-learning, community service, volunteering, and reflection). For example, some researchers limit their studies to service-learning experiences that occur in credit-bearing courses; others include co-curricular service in their definitions. This disparity leads to conclusions that may not be compatible. Lack of clarity and specification of conceptual and procedural aspects of the research can severely limit the value of information collected.

- *Measures are mostly of the self-report type*: Most service-learning research that involves student measures utilizes tools that are based on self-report (e.g., students self-report that they learned a great deal about diversity in a service-learning class). Although self-report instruments can be useful, they also have limitations (see Bringle et al., 2004; Steinke & Buresh, 2002), including that they may be influenced by social desirability response sets, they may not correspond to behavior, they may not accurately reflect processes that determined outcomes, and they may be affected by inaccurate or biased memories A few studies have utilized other types of tools such as behavioral ratings by an external observer and coding of student products (e.g., Osborne, Hammerich, & Hensley, 1998; Ash et al., 2005.)

Qualitative Designs

Qualitative research focuses on analysis of documents, artifacts, words, pictures, and other non-numerical data. Often the researcher has direct contact with the persons being studied in the research and may even be a participant observer; therefore, the researcher's insights and experiences may form an important part of the investigation. The approach is descriptive, interpretative, subjective, and inductive in nature. Qualitative studies are characterized by three design strategies (Denzin & Lincoln, 2005; Guba & Lincoln, 1989, 2005; Kiely & Hartman, in press; Patton, 2002; Schwandt, 2001), described below.

Naturalistic Inquiry

In qualitative research, the investigator observes what is taking place without attempting to manipulate or control the phenomena of interest. Naturalistic inquiry occurs in real world settings (i.e., not in a laboratory or artificial setting), and events unfold naturally without interventions or controls on the setting or a course predetermined by the researcher. The investigator does not attempt to put constraints on study conditions or outcomes. This approach is characterized by openness to whatever emerges from the study, and requires that the researcher pay attention to process and situation dynamics that may change over the course of the study.

Emergent Design Flexibility

Qualitative research design involves an iterative process because the investigator uses a discovery-type approach in determining instruments to be used and data to be gathered. That is, the researcher may start out looking at a certain type of document, noticing particular patterns and themes, then as events unfold *in the midst of the study* decide to switch to oral interviews to get a different perspective on the themes. This iterative process might be repeated several times before the researcher creates a synthesis and interprets the data. Thus, the design of qualitative research is open-ended in nature. While the researcher initially specifies an outline for what is to be accomplished and plans for observations such as initial interview questions, the approach may be altered as the situation changes and new paths of inquiry emerge.

Purposeful Sampling

Contrary to quantitative research, which might involve deliberate attempts to obtain representative or random samples, qualitative researchers do not attempt to gain a representative sample of a population under study. The focus in qualitative research is on gathering data from "information rich" cases that can be studied to provide an in-depth understanding of the phenomena of interest. Thus, qualitative researchers *purposefully* sample the specific documents, artifacts, people, and programs that are illuminative and that they believe will provide a great deal of information related to the purpose of the research.

Often this type of research results in small sample sizes or even ***single case studies*** (N=1). As noted above, one of the common problems in service-learning research is small sample sizes that limit the generalizability of the data and inferences, making it difficult to have confidence in the results and generalize to other situations and samples. Too often researchers describe a single case study with limited data, such as a single service-learning course or program. Although this may prove helpful to others trying to conceive a new course or program, generally it does not lead to new cumulative understandings or contribute to the research knowledge base on service-learning. Qualitative researchers are less concerned with these issues,

often valuing a few meaningful cases for study, rather than large numbers of less meaningful cases. Nevertheless, we recommend triangulation through the use of multiple measures and conducting cross-case (multi-case) analysis to increase understanding and generalizability. Cross-case analysis involves making comparisons between cases, analyzing relationships, and hypothesizing about causes and consequences (Patton, 2002).

Quantitative Research Design

As indicated above, quantitative research focuses on analysis of numeric data. The approach often follows particular scientific methods (e.g., design, sampling, measurement). Quantitative research can be classified into three types shown in Table 4 (Trochim, 2006).

Table 4. Quantitative Research Designs

Research Design	Non-Experiment	Quasi-Experiment	Experiment
Random Assignment of Subjects to Groups	No	No	Yes
Control Group or Multiple Waves of Measurement	No	Yes	Yes

Non-experimental designs

Non-experimental designs do not involve random assignment of subjects to groups, nor is there a control or comparison group. Non-experimental designs also do not involve multiple waves of measurement. This type of design is very useful for descriptive research questions such as:

- What percentage of students is involved in community service?
- Do male students have different attitudes than females about the need for social service agencies?
- How many faculty members have taught a service-learning course in the past three years?

The simplest, very common form of non-experiment is a *one-shot survey*. For example, a researcher might conduct a survey of opinions about community activism. In a variation on this, a researcher might measure attitudes at the end of a semester in a service-learning course. This design (called the *post-test only, single group design*, Campbell & Stanley, 1963) lacks a comparison, and therefore the ability to conclude that the outcome was the result of the service-learning experience.

Correlational research designs evaluate the nature and degree of association between two naturally occurring variables. The ***correlation coefficient*** is a statistical summary of the nature of the inferred association between two constructs that have been operationalized as variables. The correlation coefficient contains two pieces of information (a) a number, which summarizes the degree to which the two variables are linearly associated; and (b) a sign, which summarizes the nature or direction of the relationship. The numeric value of a correlation coefficient can range from +1.0 to -1.0. Larger absolute values indicate greater linear association; numbers close to zero indicate no linear relationships. A positive sign indicates that higher values on one variable are associated with higher values on the other variable; a negative sign indicates an inverse relationship between the variables such that higher values on one variable are associated with lower values on the other variable.

Causal inferences are very difficult to make from a single correlation because the correlation does not assist in determining the direction of causality. For example, a positive correlation between volunteering and self-esteem indicates that more volunteering is associated with higher self-esteem. However, the correlation does not differentiate among at least three possibilities, (a) that volunteering affects self-esteem; (b) that self-esteem promotes volunteering; or (c) that a third variable (e.g., self-efficacy) is responsible for the correlation between self-esteem and volunteering.

Experimental designs

In contrast to correlational methods that assess the patterns between naturally occurring variables, ***experiments*** manipulate a variable, the ***independent variable***, and see what consequence that manipulation has on another variable, the ***dependent variable***. Not all experimental designs are equally good at allowing the researcher to make causal inferences. An outline of experimental designs is presented below. Note that this section is only intended to be an introduction to the topic. For more specific information on experimental research design, the reader should consult a research methodology text (e.g., Campbell & Stanley, 1966; Cook & Campbell, 1979; Cozby, 2009; Kerlinger, 1986). Consultation with experienced research colleagues is also helpful. Some online resources on design are listed in the appendices of this document.

The strongest research design in terms of drawing cause-and-effect conclusions (internal validity) is the ***randomized*** or ***true experiment***. In this "gold standard" of quantitative research designs, subjects are randomly assigned to different groups or treatments in the study. Traditionally these groups of subjects are referred to as the *experimental* or *treatment group(s)* (e.g., students in a service-learning course) and the *comparison* or *control group(s)* (e.g., students in a traditional course). Note that random *assignment* of subjects to a group in an experiment is different from the random *selection* of subjects to be involved in the study. Random assignment makes it unlikely that the treatment and control groups differ significantly at the beginning of a study on any relevant variable, and increases the likelihood that differences on

the dependent variable result from differences on the independent variable (treated group vs. control group). Random assignment controls for self-selection and pre-existing differences between groups; random selection or sampling is relevant to the generalizability or external validity of the research.

There are a variety of designs that utilize random assignment of subjects, but true experimental studies are relatively rare in service-learning research, as in most educational research. This is because it is usually difficult, especially in higher education settings, to randomly assign students to service-learning versus traditional courses or to different levels of a variable in the instruction. Nevertheless, the U.S. Department of Education has proposed that all research use random assignment so that education practice can be based on research with internal validity (www2.ed.gov/rschstat/eval/resources/randomqa.html). A close approximation of random assignment occurs when students are not aware that some sections of a course will be service-learning and some will not be service-learning when they register for courses (Markus, Howard, & King, 1993; Osborne et al., 1998). Also, there may be opportunities to randomly assign students to different conditions in service-learning classes (e.g., students are randomly assigned to (a) written reflection or (b) reflection through group discussion).

Quasi-experimental designs

Like experimental designs, quasi-experimental designs involve the manipulation of an independent variable to examine the consequence of that variable on another (dependent) variable. The key difference between experimental and quasi-experimental designs is that the latter do not involve random assignment of subjects to groups. A large portion of past quantitative research on service-learning involves quasi-experimental design.

We do not intend to comprehensively cover all quasi-experimental designs in this primer; instead we will discuss some designs commonly seen in service-learning research. For more advanced information, or for information on other designs not discussed here, we recommend that the reader consult a graduate-level research methodology text (e.g., Campbell & Stanley, 1966; Cook & Campbell, 1979; Cozby, 2009; Kerlinger, 1986). Consultation with experienced research colleagues is also helpful. In addition, some online resources are listed in the appendices of this document.

One aspect of designing a study relates to temporal arrangements. Some researchers are interested in the developmental aspects of service-learning, or in the effects of service-learning over time. For example, they may be interested in the question of whether involvement in volunteer service during high school leads to increased involvement in service during and after college. There are two approaches to designing research to answer these types of questions. In a *cross-sectional design* the researcher gathers data from several different groups of subjects at approximately the same point in time. For example, a researcher might choose to conduct interviews with groups of college freshmen, juniors, graduating seniors, and alumni. *Longitudinal studies* (sometimes also called *time series designs*) involve gathering information

about one group of people at several different points in time. Astin, Sax, and Avalos (1999), for example, collected survey data from entering freshmen in 1985, then surveyed the same group of students four years later in 1989, and again to the now-alumni in 1994-95. Longitudinal studies are extremely valuable sources of information for studying long-term consequences of service-learning, but they are rare in service-learning research because of the practical, technical, and financial difficulties in following a group of people over time.

Other researchers focus their interest on questions that do not relate to developmental issues or impact over a long period of time. In fact, many if not most service-learning studies are limited to one semester or sometimes one year in length. A common strategy is to give an attitude measure to students in a service-learning course at the beginning and end of a semester. This ***pre-test, post-test single group design*** examines the difference between pre- and post-test scores for one group of students. Unfortunately, there is no assurance that the difference in pre-test and post-test scores is due to what took place in the service-learning class. The difference in attitudes could be attributable to other events in the students' lives (history), natural growth changes (maturation), dropout of the least motivated students during the course (mortality), or carryover effects from the pre-test to the post-test (testing).

Another experimental design is the ***post-test only, static groups design***[1], which compares the outcomes of a pre-existing treated group to the outcomes of a pre-existing untreated group. Using this design, an instructor could give an attitude scale at the end of the semester to a service-learning section of the course and also to a section that did not contain service-learning. This design suffers from the limitation that it is not possible to conclude that the difference on the dependent variable, attitudes, is due to the difference in instruction because it is not known if the two groups were equivalent in their attitudes at the beginning of the semester.

An alternative arrangement, the ***nonequivalent (or untreated) control group design with pre- and post-test***, is to give a pre-test and a post-test to both a service-learning section of a course and to a traditional section that does not include a service component. In this design the researcher can evaluate whether or not the two groups were equivalent at the beginning of the semester, but only on the measured variables. A second step is to examine the pattern of changes between the two groups across the semester.

The biggest problem with the nonequivalent groups design is ***self-selection bias,*** described above in the section "Common Problems in Service-Learning Research." Frequently in higher education, and sometimes in high school settings, service-learning courses are optional for graduation, and/or service is an optional component of a particular course. That is, students must select or opt to be in the class and to participate in service. The result is that students are non-randomly assigned to the treatment group (service-learning course) and thus there is non-random assignment of students to groups. There are likely to be many differences between students who

---

[1] The pre-test, post-test single group design and the post-test only static groups design are sometimes classified as non-experiments or pre-experiments (Campbell & Stanley, 1966) because the designs generally do not permit reasonable causal inferences. Later authors (Cook & Campbell, 1979; Trochim, 2006) include these designs in the category of quasi-experiments.

choose to be involved in service-learning classes and those who do not (Eyler & Giles, 1999). Even with a pre-test to compare equivalence of groups at the beginning of the study, a researcher could never completely eliminate the possibility that there are differences on other, unpre-tested variables, or that post-test differences are due to inherent differences in the groups, rather than differences in the educational intervention. Sometimes researchers use multiple measures pre- and post-treatment to help assess whether groups are equivalent on several relevant variables; statistical procedures (i.e., analysis of covariance) also can help control for differences between treatment and non-treatment groups, but only for measures that are obtained prior to the educational intervention. Of course, the best solution is random assignment of students to groups, which makes this an experimental design, rather than a quasi-experimental one.

A common variation of the nonequivalent groups design occurs when students in two sections (one including a service component and one not) of a course are being compared, but the two sections are taught by different instructors. This creates a problem in interpretation because one cannot infer that post-test differences in scores are due to the style of pedagogy (service-learning) rather than other differences between instructors. Another variation is to compare two sections of the same course, one involving service and one not, but taught in different semesters. In this case it is possible that differences in post-test scores are due to events extraneous to the study, which happened during one semester but not the other. In sum, it is important for the researcher to be aware of potential pitfalls of any research design and to take these into account when drawing conclusions from the study.

Chapter 3
Measurement in Service-Learning Research

An important component of designing a research project is deciding on the measures to be used. This chapter focuses on types of assessment tools, characteristics of good instruments, and other considerations for choosing measurement approaches.

Common Types of Assessment Tools in Service-Learning Research

Several types of measurement procedures are common in research on service-learning: surveys and questionnaires, rating scales, interviews, focus groups, observational checklists, and rubrics for content analysis of student reflections. Instruments that were designed for course or program evaluation purposes can sometimes be adapted for research purposes. Each of the common research measurement tools is described below.

Surveys and Questionnaires

One of the most commonly-used research tools is the survey (also called a questionnaire). Surveys may be conducted in several ways: face-to-face, by telephone, by email, on the internet, or on paper. Surveys frequently incorporate *rating scales*, discussed below.

Many surveys that are used in service-learning research are *self-report* measures. In this type of survey, respondents report on their own attitudes, opinions, behaviors, behavioral intentions, feelings, or beliefs. They may ask about the occurrence of an event (e.g., "Were you nervous on your first day of service?"), the intensity (e.g., "How nervous?"), frequency (e.g., "How often did you tutor at the service site?"), and the degree of endorsement (e.g., "I was extremely nervous." "Strongly Agree," "Agree," "Disagree," "Strongly Disagree"), or about the likelihood (e.g., "How likely is it that you will be nervous at your next visit?" 'Very Likely' to 'Very Unlikely'"). Self-report surveys are very useful for many research purposes because they obtain information directly from the respondent; however, researchers should keep in mind that self-report measures have several important drawbacks. One disadvantage is that they are subjective and may not coincide with ratings given by other sources of information (e.g., the instructor, an outside observer, another student, a staff member). Another drawback is that they may be subject to *social desirability bias* (the tendency for a person to give responses that are normative and present oneself in a good light).

Surveys and questionnaires may be composed of scales, which are intentionally designed coherent measures of a construct (e.g., trait, attribute) that combine multiple indicators. Although most *questionnaires* are composed of several scales, typically each individual *scale* is a multiple-item measure of only one construct. As such, a scale should display qualities consistent with the assumptions of being unidimensional and measuring only one construct (single factor structure, high coefficient alpha). Bringle, Phillips, and Hudson (2004) present a discussion of

the characteristics of good scales, and they provide examples of scales that can be used in service-learning research.

   Rating scales in surveys: Many surveys incorporate rating scales. Probably the most common format is known as a **Likert-type response format**, in which the study participant chooses an option from a *list indicating his or her level of agreement or disagreement with a statement*. For example:

   I do not know what I would do without my cell phone.
   1. Strongly Disagree
   2. Disagree
   3. Agree
   4. Strongly Agree

Many variations on this type of rating item are possible, including the presence or absence of a neutral point and the number of choices (e.g., 4, 5, 6, 7). A less-commonly used scale format asks participants to *give their opinion* on some issue, experience, or product using a subjective rating scale. For example:

   Rate the quality of the movie *Jaws* on a scale of 1 to 10, with 1 being low and 10 being high.

Yet another scale format requests participants to *rank statements in order of preference or agreement.* An example of this is the following:

   Place a number next to each of the following to indicate your preference, with 1 being your first choice, 2 being your second choice, and so on.
   ____ Working at a food pantry
   ____ Tutoring children in an urban school
   ____ Cleaning the grounds of a park
   ____ Painting walls in a community center

   Checklists in surveys: Checklists are another common element of surveys. Frequently these are seen with instructions such as "select one" or "check all that apply." For example:

   Please indicate which activities you have participated in during the past year (check all that apply):
   ___ Participating in community service through a course (service-learning)
   ___ Volunteering for a service activity through campus, such as United Way Day of Caring
   ___ Participating in a public debate, working on a political campaign, or assisting with voter
       registration
   ___ Community involvement through a campus organization or club
   ___ Community service or involvement as part of a financial aid package

___ Service through another organization not connected to the university

A researcher may wish to have subjects indicate how often they have participated in particular activities, for example:

Please indicate how often you have participated in the following in the past year:
          1 = None/Never
          2 = Once each school year
          3 = Once or twice each semester
          4 = About once a month
          5 = Nearly every week or more
___ Participating in community service through a course (service-learning)
___ Volunteering for a service activity through campus, such as United Way Day of Caring
___ Participating in a public debate, working on a political campaign, or assisting with voter registration
___ Community involvement through a campus organization or club
___ Community service or involvement as part of a financial aid package
___ Service through another organization not connected to the university

Interviews and Focus Groups

An ***interview*** is another research tool that is especially useful at the exploration stage, or for qualitative research. Interviews can be conducted either in person or by telephone. They are similar to surveys, but are often used to assess information in more depth than would be possible in a survey. Interviewers usually have a protocol for asking questions and obtaining responses. Interview questions maybe open-ended (i.e., content determined by the interviewer) or structured (pre-determined content and order). Also, the responses can be open-ended (the respondent is free to say anything) or close-ended (pre-determined responses categories are chosen by the interviewer). Furthermore, interviews might be recorded (e.g., audio-taped or videotaped) for later analysis. Taped interviews can be transcribed for data analysis by judges.

Focus groups are interviews that are conducted in a group format. One of the biggest advantages of focus groups is that participants can interact and build on comments from each other (which may be offset by uneven participation). Another advantage is in saving time by conducting the interviews in groups rather than one-on-one. Shortcomings of focus groups include that the group format may suppress information from some respondents and the qualitative data analysis may be as time-intensive as analysis of interviews. Another disadvantage of focus groups is that participants may not have time or feel free to make completely honest comments in front of others. Focus groups also may not be as useful as interviews for getting in-depth information about a particular individual's experiences.

Interviews and focus group tapes generally must be transcribed for data analysis, often in the form of content analysis. ***Content analysis*** is described in the section "Reflection" below. In general, interviews are more expensive to conduct than surveys and questionnaires. There are

expenses for training the interviewers, getting the respondents and the interviewer together, and the time for the interview. In addition, there is the risk that the interviewer's characteristics (e.g., gender, race, age) and paralinguistic cues will influence the respondent's answers. These shortcomings are attenuated or eliminated in written questionnaires.

Observation Rating Scales and Checklists

Sometimes a researcher may wish to have observations of behaviors made in a classroom or at a service site. This is especially useful for providing corroborative evidence to supplement information that students have supplied through surveys or reflections, or for corroborating information given by others (peers, community partners) at the service site.

One way to record observations is to keep a journal or log. This assessment method usually would be used to triangulate with other research data. To record observations in a more quantitative format, a researcher might choose to use a rating scale or checklist. An *observational rating scale* usually instructs the observer to rate the frequency, quality, or other characteristic of the behavior being observed, such as:

Number of times the tutor established eye contact with student X:
      0   1-2   3-5   10
Quality of nursing student's interactions with community health center staff:
     Low   Medium Low   Medium High   High

Behaviorally anchored rating scales detail the particular dimensions of action that a rater is to look for and requires the rater to determine the absence or frequency of behaviors that are indicative of the dimensions.

In an *observational checklist* or *inventory* an observer would make a checkmark on a list when a behavior was observed, for example:

√   Tutor established eye contact with student
        Tutor smiled at student
√   Tutor touched student in appropriate manner
√   Tutor used language appropriate to the age and abilities of the student
3_   Total number of check marks

Document Review

A research project may require review of documents such as reflection products, course syllabi, faculty journals, meeting minutes, strategic plans, annual reports, or mission statements. These artifacts provide a rich source of information about programs and organizations. One limitation of documents and records is that they may be incomplete, inaccurate, or may vary in quality and quantity (Patton, 2002). Document review is usually associated with the qualitative

approach to research, but depending on the research question, the researcher might utilize a rating scale, checklist, or rubric to summarize qualitative data from documents. Review of journals and reflections might also involve qualitative content analysis (described below). Gelmon, Holland, Driscoll, Spring, and Kerrigan (2001) provide examples of document review for service-learning faculty and institutional questions. Although their examples are designed for program evaluation purposes, they can be adapted for research purposes.

Reflection Products: Content Analysis and Rubrics

One of the most common tools used for service-learning assessment is student (or faculty) reflection products. Reflection can take many forms such as case studies, journals, portfolios, papers, discussion, presentations, and interviews. (For a discussion of reflection activities and how to structure reflection to enhance student learning, see www.compact.org/disciplines/reflection/) For research purposes, reflections are typically analyzed by one of two methods: content analysis, or rubrics.

Content analysis is a standard social science methodology for studying the content of communication. "Generally…content analysis is used to refer to any qualitative data reduction and sense-making effort that takes a volume of qualitative material and attempts to identify core consistencies and meanings…often called patterns or themes" (Patton, 2002, p. 453). In service-learning research, much content analysis is informal in nature. In this technique, researchers develop a series of themes, categories, or coding frames. The process of discovering patterns, categories, or themes in the data is called *inductive analysis* or *open coding*. If a framework already exists, the process is called *deductive analysis* (Patton, 2002). Reflection products are coded against the categories, leading to deductions about common themes, issues processes, or ideas expressed, as well as student development along academic, social, or civic dimensions.

More formal content analysis is used when there are large amounts of data to be analyzed (see Eyler and Giles, 1999, for an example of theory-guided content analysis). Software programs have been developed, such as NVivo, to assist in content analysis. These programs are sometimes called "computer-assisted qualitative data analysis software", or CAQDAS. These programs code narratives based on keywords, themes, key phrases, or other salient features. The most widely available software programs are used on text materials, but programs such as NVivo can also be used to analyze audio, video, and other media. For more information on content analysis see:

- CAQDAS Comparison
  (www.restore.ac.uk/lboro/research/software/caqdas_comparison.php)
- Choosing a CAQDAS Package
  (caqdas.soc.surrey.ac.uk/PDF/2009ChoosingaCAQDASPackage.pdf)
- Wikipedia: Content Analysis (en.wikipedia.org/wiki/Content_analysis)

Rubrics offer another way to analyze reflection products or other artifacts. A rubric is a scoring tool for subjective assessments, allowing for more standardized evaluation of products based on specified criteria. Rubrics can be either holistic (one-dimensional) or analytic, providing ratings along several dimensions. Table 3 presents an example of an analytic rubric developed by the IUPUI Center for Service and Learning. Rubrics usually occur in the form of a matrix, with the following characteristics:

- Traits or dimensions that serve as the basis for judging products
- Definitions or examples to illustrate the traits or dimensions
- A scale of values on which to rate the traits
- Standards or examples for each performance level

Researchers should have multiple people providing ratings of each reflection or artifact, and should establish the ***inter-rater reliability*** of any rubrics used in an investigation. Inter-rater reliability is the degree to which different observers or raters give consistent scores using the same instrument, rating scale, or rubric. Knowing the inter-rater reliability and using multiple raters helps to establish the credibility of the rubric being used and helps the investigator feel confidence in the results and conclusions coming from the research.

A good source for pre-made, editable rubrics is RubiStar (rubistar.4teachers.org). Other rubrics for evaluating student reflections in service-learning courses are available online:
- Rubric for Coding Problem-Solving Analysis Protocol (P-SAP) (www.ncsu.edu/assessment/resources/Rubric.pdf)
- Rubric to Assess Service Learning Reflection Papers (www.compact.org/wp-content/uploads/resources/downloads/SL_Reflect_Form2.pdf)

Pre-existing Data Sources

Researchers also can conduct ***secondary data analysis*** on data that has been collected by another researcher. Research using pre-existing data should be guided by theory, focus on clear research questions or hypotheses, and be consistent with the constraints of the data (sampling, subject population, measurement, design). Although so far this technique has not been used extensively in service-learning research, NSLC has begun to compile sets of data for secondary data analysis. Two existing sources of data that are particularly relevant for service-learning researchers are:

- Higher Education Research Institute (HERI) at UCLA (gseis.ucla.edu/heri)
- Center for Information and Research on Civic Learning and Engagement (CIRCLE) Data Sets: (www.civicyouth.org/ResearchTopics/research-products-cat/data-sets/)
- NSLC will continue to add data sets to the Conducting Research page (servicelearning.gov/service-learning-research-tools/conducting-research)

Table 3. Sample scoring rubric for a student journal

Levels	Criteria
Reflective practitioner	Clarity: The language is clear and expressive. The reader can create a mental picture of the situation being described. Abstract concepts are explained accurately. Explanation of concepts makes sense to an uninformed reader. Relevance: The learning experience being reflected upon is relevant and meaningful to the student and course learning goals. Analysis: The reflection moves beyond simple description of the experience to an analysis of how the experience contributed to student understanding of self, others, and/or course concepts. Analysis has both breadth (incorporation of multiple perspectives) and depth (premises and claims supported by evidence). Interconnections: The reflection demonstrates connections between the experience and material from other courses, past experience, and/or personal goals. Self-criticism: The reflection demonstrates ability of the student to question biases, stereotypes, preconceptions, and/or assumptions and define new modes of thinking as a result.
Aware practitioner	Clarity: Minor, infrequent lapses in clarity and accuracy. Relevance: The learning experience being reflected upon is relevant and meaningful to the student and course learning goals. Analysis: The reflection demonstrates student's attempts to analyze the experience but analysis lacks depth and breadth. Interconnections: The reflection demonstrates connections between the experience and material from other courses, past experience, and/or personal goals. Self-criticism: The reflection demonstrates ability of the student to question biases, stereotypes, and preconceptions.
Reflection novice	Clarity: There are frequent lapses in clarity and accuracy. Relevance: Student makes attempts to demonstrate relevance, but the relevance is unclear to the reader. Analysis: Student makes attempts at applying the learning experience to understanding of self, others, and/or course concepts but fails to demonstrate depth and breadth of analysis. Interconnections: There is little to no attempt to demonstrate connections between the learning experience and previous personal and/or learning experiences. Self-criticism: There is some attempt at self-criticism, but the self-reflection fails to demonstrate a new awareness of personal biases, etc.
Unacceptable	Clarity: Language is unclear and confusing throughout. Concepts are either not discussed or are presented inaccurately. Relevance: Most of the reflection is irrelevant to student and/or course learning goals. Analysis: Reflection does not move beyond description of the learning experience(s). Interconnection: No attempt to demonstrate connections to previous learning or experience. Self-criticism: Not attempt at self-criticism.

Developed by Stephen Jones, IUPUI Center on Service and Learning

Other sources of data for secondary analysis can be found at the following websites:

- National Center for Education Statistics (NCES)—a branch of the US Dept. of Education (nces.ed.gov)
- National Opinion Research Center (NORC) at the University of Chicago (www.norc.uchicago.edu/channels)
- Survey Documentation and Analysis page at the University of California, Berkeley (sda.berkeley.edu/archive.htm) :
- Institute for Social Research (University of Michigan) (www.isr.umich.edu/home/projects)
- Inter-University Consortium for Political and Social Research (University of Michigan) (www.icpsr.umich.edu/icpsrweb/ICPSR/access/subject.jsp)
- Scientific Research on the Internet (University of Maryland, College Park) (www.webuse.umd.edu)
- Minnesota Population Center at the University of Minnesota (www.ipums.umn.edu)
- National Bureau of Economic Research (www.nber.org)
- Data Access Tools from the U.S. Census Bureau (www.census.gov/main/www/access.html)
- IUPUI Center on Philanthropy Panel Study (www.philanthropy.iupui.edu/Research/giving_fundraising_research.aspx#COPPS)

Many colleges and universities have campus-wide data from surveys of students, faculty, and staff that can be used for comparison purposes. For example, the IUPUI page on Student, Staff and Faculty Surveys (www.planning.iupui.edu/95.html) includes results from the National Survey of Student Engagement (NSSE) and Faculty Survey of Student Engagement (FSSE).

Characteristics of Good Measurement Instruments

Procedures for measuring attributes can be judged on a variety of merits. These include practical well as technical issues. All measurement procedures, whether qualitative or quantitative, have strengths and weaknesses—no one procedure is perfect for every task. In order to improve a study it is frequently prudent for an investigator to use multiple measurement tools and triangulate the results.

Practical Issues

Some of the practical issues that need to be considered for each tool include:

- Cost
- Availability

- Training required
- Ease of administration, scoring, analysis
- Time and effort required for respondents to complete the measure
- Completeness of the data gathered
- Potential sources of bias
- Relevance to research question

Along with the practical issues, quantitative measurement procedures (especially surveys, tests, and scales) may be judged on the ***technical characteristics*** or ***psychometric properties*** of the instruments. There are two major categories of psychometric properties—reliability and validity—both of which are important for good quantitative research instruments. The following description is a general outline of the major forms of reliability and validity. For more specific information the reader is urged to consult a good text on psychometrics (e.g., Furr and Bacharach, 2008).

Consistency (Reliability)

A good measure of some entity is expected to produce consistent scores. A procedures' reliability is estimated using a coefficient (i.e., a numerical summary). For purposes of service-learning research, the major types of coefficients include:

- ***Temporal consistency***: the ability of an instrument to give accurate scores of the same entity from one time to another. Also known as ***test-retest reliability***, it uses the correlation coefficient between the two administrations of the same scale.

- ***Coherence***: the consistency of items within a scale. ***Internal consistency reliability*** estimates the consistency among all items in the instrument (typically measured using Cronbach's coefficient alpha). According to Nunnally (1967), coefficient alpha is an estimate of the correlation between the scale and a hypothetical alternative form of the scale of the same length. Alternatively, it is an estimate of the correlation of the scale with the construct's true score. An important principle that is related to coefficient alpha is that, other things being equal (e.g., item quality), the more items a scale contains, the more reliable, coherent, and error free it will be.

- ***Scoring agreement***: the degree to which different observers or raters give consistent scores using the same instrument, rating scale, or rubric. Also called ***inter-rater reliability***, it is a particularly important consideration when using rubrics. Knowing the inter-rater reliability helps to establish the credibility of the rubric being used and helps the investigator feel confidence in the results and conclusions coming from the research.

For more information about reliability please refer to Types of Reliability (www.socialresearchmethods.net/kb/reltypes.php) and Reliability Analysis (faculty.chass.ncsu.edu/garson/PA765/reliab.htm).

Meaningfulness (Validity)

A valid measurement tool or procedure does a good job of measuring the concept that it purports to measure. Validity of an instrument only applies to a *specific purpose with a specific group of people*. For example, a scale is not considered simply "valid" or "invalid"—but it might be considered valid for measuring social responsibility outcomes with college freshmen, but not knowledge of the nonprofit sector among professionals. Below are three main classes of validity, each having several subtypes.

- *Construct validity*: The theoretical concept (e.g., intelligence, moral development, content knowledge) that is being measured is called the **construct.** Construct validity establishes that the procedure or instrument is measuring the desired construct because the operationalization (e.g., scores on the scale) conforms to theoretical predictions. This is the most important form of validity, because it subsumes other forms of validity.
  - *Convergent validity***:** correlation of scores on an instrument with other variables or scores that should theoretically be similar. For example, two measures of social responsibility should yield similar scores and therefore be highly correlated.
  - *Discriminate validity***:** Comparison of scores on an instrument with other variables or scores from which it should theoretically differ. For example, a measure of verbal ability should not be highly correlated with artistic skills.
  - *Factor structure***:** Factor analysis provides an empirical examination of the internal consistency of an instrument. The items that are theoretically supposed to be measuring one concept (i.e., a subscale) should correlate highly with each other and all load on the same factor, but have low correlations with items measuring a theoretically different concept (an orthogonal or independent factor). In some cases, the theoretical construct might have multiple dimensions and the factor structure will not be unidimensional, but the factor structure should correspond to the theoretical structure.

- *Content validity:* Establishes that the instrument includes items that are judged to be representative of a clearly delineated content domain. For example, the IUPUI Center for Service and Learning established a framework of knowledge, skills, and dispositions of a civic-minded graduate (Bringle & Steinberg, in press). They used this conceptual framework to develop items for an instrument, the Civic-Minded Graduate Scale. Content validity can be assessed by the degree to which the scale items are a representative sample of a clearly defined conceptual domain according to the evaluation of independent reviewers.

- *Face validity*: A subjective judgment about whether or not, on the "face of it," the tool seems to be measuring what it intended to measure.

- **Criterion-related validity**: The degree to which an instrument is associated with a criterion that is implicated by the theory of the construct.
  - *Concurrent validity***:** Comparison of scores on some instrument with concurrent scores on another criterion (e.g., behavioral index, independent assessment of knowledge). If the scale and the criterion are theoretically related in some manner, the scores should reflect the theorized relationship. For example, a measure of verbal intelligence should be highly correlated with a reading achievement test given at the same time, because theoretically reading skill is related to verbal intelligence.
  - *Predictive validity***:** Comparison of scores on an instrument with some future criterion (e.g., behavior). The instrument's scores should do a reasonable job of predicting the future performance. For example, scores on a social responsibility scale would be expected to be fairly good predictor of future post-graduation civic involvement (e.g., voting, volunteering).

For more specific information about test validity, see the following web pages:
- Validity: Statnotes, from North Carolina State University, Public Administration Program (faculty.chass.ncsu.edu/garson/PA765/validity.htm)
- Instrument Validity - Educational Research - Del Siegle (www.gifted.uconn.edu/siegle/research/Instrument%20Reliability%20and%20Validity/Validity.htm)
- Validity (statistics) – Wikipedia (en.wikipedia.org/wiki/Validity_%28statistics%29)

For information on the relationship between reliability and validity, see Reliability & Validity (www.socialresearchmethods.net/kb/relandval.php).

Web Survey Tools

Surveys can be conducted in a variety of ways, by mail, telephone, email, online, or face-to-face. Each of these methods has advantages and disadvantages. Online surveys are becoming increasingly popular because of their convenience and low cost. Negative aspects of online surveys include potential issues of confidentiality, non-random sampling errors, and low response rates. The following are resources about creating online surveys:
- Creating Online Surveys for Your Non-profit Organization (www.tsne.org/site/c.ghLUK3PCLoF/b.3524717/k.767D/Articles__Online_Surveys.htm)
- Survey Research (www.socialresearchmethods.net/kb/survey.php)

There are many online tools for conducting surveys. The following list includes some of the most popular and readily available.

- Insite Survey Systems (www.insitesystems.com/surveys/)
- Respondus (www.respondus.com/)
- Checkbox (www.checkbox.com/features/overview.aspx)
- Zoomerang (www.zoomerang.com/)
- SurveyGizmo (www.surveygizmo.com/)
- SurveyMonkey (www.surveymonkey.com/)
- QuestionPro (www.questionpro.com/)
- SurveyGold (www.surveygold.com/)
- eSurveysPro (www.esurveyspro.com/)
- SuperSurvey (www.supersurvey.com/)

Using Existing Instruments versus Modifying versus Creating

An important decision point when planning research is whether to use existing tools, adapt from those that have already been developed, or create new tools to fit a specific purpose. Some of the pros and cons of each option are listed below in Table 4.

Table 4. Deciding on Instruments

Using Existing Instruments

Pros:	Cons:
Can be incorporated into research quickly	May not "fit" research question exactly
Often prepared by professional expert	May require training for administration, scoring, or analysis
May have norms for comparison purposes	May incur cost to purchase, score, or analyze
May have known reliability and validity indicators	May be too long for the purposes at hand, take too much time to complete
Can build on existing knowledge base using the same instrument	
May be keyed to content or proficiency standards	

Adapting Existing Tools
(E.g., using portions of an instrument, small wording changes, changing the time frame)

Pros:	Cons:
Can be modified to suit research question or service-learning context	Changing a known quantity into something unknown
Most of the work of creating the tool has been completed	Previous reliability and validity indicators may no longer apply
May be able to compare results with previous results or norms (but only on relevant items or subscales)	
Using only a portion of an existing instrument may lessen completion time, and thus increase response rate	

Developing New Instruments

Pros:	Cons:
Can develop instrument to fit specific need	Requires time, effort, resources, expertise
Instrument itself may comprise a significant contribution to the field of research	Requires knowledge of scale development procedures
	Runs risk that instrument will not be reliable or valid for purpose at hand

Chapter 4
Ethical Issues in Service-Learning Research

Chapdelaine, Ruiz, Warchal, and Wells (2005) posit that "despite a tremendous amount of writing and research on the mechanics of doing service-learning . . . there is a paucity of literature on the ethical challenges involved with this pedagogical undertaking, especially in higher education" (p. xi). Because ethical considerations in service-learning research may potentially be overlooked by researchers, those who are (or will be) involved in research on service-learning course outcomes need to be aware of and responsive to ethical considerations when designing, implementing, or subsequently researching a service-learning course. Like most other social sciences, research in service-learning involves human subjects (e.g., students, faculty, community-based agency staff, clients); therefore, it is governed by the same ethical codes as any other research that also involves the participation of humans as subjects.

Historical Overview

Schneider (n.d.) states that "the final developments that produced our current method of protecting human research subjects were the result of historical events in the twentieth century" (www.iupui.edu/~histwhs/G504.dir/irbhist.html). These included such high profile ethically problematic research cases as the Tuskegee Syphilis Study (Brandt, 1978; Rothman, 1982), and the research in Nazi concentration camps during World War II (Schneider, n.d.). The major outcome of the Nazi atrocities was the adoption of the Nuremberg Code (see *Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10,* 1949), while the Tuskegee Study prompted the Belmont Report in order to protect human subjects (see The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Both of these reports produced many of guidelines and regulations that now guide researchers who conduct studies that involve using human research participants. These include (but are not limited to): minimal risk to the subject; obtaining the informed consent of the subject; the right of the subject to withdraw from the study at any point; any benefits from the research must be greater than its associated risks; the study must be conducted by qualified individuals; and the study must be able to be stopped at any point during the course of the research.

Basic Ethical Principles

The Belmont Report lists three major ethical principles: Respect for Persons, Beneficence, and Justice (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Beauchamp and Childress (2001) expand upon the Belmont Report's principles by identifying four guiding ethical principles—Respect for Autonomy, Nonmaleficence, Beneficence, and Justice.

Respect for autonomy is further defined as having "respect for the autonomous choices of persons" (Beauchamp & Childress, 2001, p. 57). This principle asserts that each person should be regarded as an autonomous being; thus they should be allowed to make their own, rational decisions whenever possible. Researchers must give subjects an opportunity for *informed consent*, indicating that they understand the nature and purpose of a study, the potential risks and benefits, and that they agree to participate in the study.

Nonmaleficence is known as the "norm of avoiding the causation of harm" (Beauchamp & Childress, 2001, p. 12). For example, a researcher must not harm any subject involved in a research study. This includes expectations of ***confidentiality*** for all data collected. In addition, human subjects must be given the ***opportunity to withdraw*** from a study at any time without penalty.

The principle of ***beneficence*** is "a group of norms for providing benefits and balancing benefits against risks and costs" (Beauchamp & Childress, 2001, p. 12). In other words, if a researcher were using humans as research subjects, the study must be of some benefit to the research participants, to science, or to society. The Institutional Review Board (IRB) is responsible for reviewing each research project to provide an independent judgment that the benefits are equal to, or outweigh, the potential risks to subjects.

Finally, the principle of ***justice*** is defined as "a group of norms for distributing benefits, risks, and costs fairly" (Beauchamp & Childress, 2001, p. 12). An example of justice is recruitment of subjects fairly from a population, with each person having an equal opportunity to be involved in a study.

Professional Codes of Ethics

Although these basic ethical principles should apply to everyone, many academic disciplines and professional organizations have their own code that members to that organization are required to uphold. For example, the Ethics Code of the American Psychological Association (APA) states that:

> Membership in the APA commits members and student affiliates to comply with the standards of the APA Ethics Code and to the rules and procedures used to enforce them. Lack of awareness or misunderstanding of an Ethical Standard is not itself a defense to a charge of unethical conduct. (American Psychological Association, 2002, p. 1)

Protection of Human Subjects (45 CFR 46)

In the United States, there are specific federal regulations that are in place to provide protection of human subjects. The United States Department of Health and Human Services Code of Federal Regulations, Title 45, Part 46 - Protection of Human Subjects, more commonly known as 45 CFR 46 (Code of Federal Regulations, 2005) provides the basics on the protection

of human subjects, but also provides guidance for the protection of pregnant women, fetuses, and neonates (Subpart B); prisoners (Subpart C); and children (Subpart D; Code of Federal Regulations, 2005).

Institutional Review Boards (IRBs)

Although research on service-learning outcomes *may* be exempt from this code because it usually involves non-medical research (see 45 CFR 46.101(b)), it is up to specific Institutional Review Boards (IRBs) to determine the exemption status (see 45 CFR 46.101(a)). Currently IRBs are the gate keepers on whether or not research in which humans will be utilized as subjects can be conducted by an investigator. According to Pimple (2006), the IRB assumes many important roles in the protection of human subjects in research.

IRB review assures that:
- risks to subjects are minimized;
- risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result;
- selection of subjects is equitable; and
- there is proper informed consent and documentation of informed consent.

In some instances, IRB review can also require that:
- the research plan makes adequate provision for monitoring the data collected to ensure the safety of subjects;
- there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data; and
- additional safeguards are included to protect the rights and welfare of any subjects likely to be vulnerable to coercion or undue influence.

Once research is initiated, IRBs have continuing responsibilities. These include:
- the conduct of continuing review at intervals appropriate to the degree of risk, and in any event, not less than once per year;
- authority to observe or have a third party observe the informed consent process and the research;
- receipt of prompt reports from investigators of any unanticipated problems involving risks to subjects or others, or any serious or continuing noncompliance with the IRB's requirements or determination, or with the regulations; and
- authority to suspend or terminate IRB approval of research that is not being conducted in accord with the IRB's requirements or that has been associated with unexpected serious harm to subjects. (Pimple, 2006, pp. 13-14)

Informed Consent

Federal regulations require that certain information *must* be provided to prospective research participants in an informed consent document. According to Pimple (2006), the basic requirements are:

- A statement that the study involves research
- An explanation of the purposes of the research.
- The expected duration of the subject's participation
- A description of the procedures to be followed
- Identification of any procedures that are experimental
- A description of any reasonably foreseeable risks or discomforts to the subject
- A description of any benefits to the subject or to others that may reasonably be expected from the research
- A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject
- A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained
- For research involving more than minimal risk, an explanation as to whether any compensation, and an explanation as to whether any treatments are available, if injury occurs and, if so, what they consist of, or where further information may be obtained
- An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights, and whom to contact in the event of research-related harm to the subject
- A statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time (pp. 21-22)

Additionally, the informed consent statement must be written at a level understandable by most of the population (not the study's sample), must include the number of other research participants in the study, the possible ways that the study can be terminated by the researcher, and the penalties (if any) for the subject voluntarily withdrawing from the study (Pimple, 2006, p. 22). There are some exceptions to the requirement for informed consent in research involving humans, such as archival research on previously-collected documents, or naturalistic observation in which researchers do not interact with participants. Parental consent may be required for studies involving minor children. In all cases an IRB must review the research protocol to determine if and how informed consent will be handled.

Sample Ethical Dilemmas

Whereas research on service-learning in higher education settings usually involves minimal risk to participants, there may be situations that involve decisions with ethical implications. These include:

- *Multi-campus studies*: Researchers involved in research on multiple campuses must consult with the IRB office of each campus. Depending on the nature of the study, the researcher *may* be required to apply for IRB approval on *each* campus.

- *International studies*: Researchers working on international service-learning courses or programs need to consult with the IRB (or equivalent) office of any non-US institutions. Investigators should also be aware of applicable laws, regulations, or norms in the country(ies) in which they are conducting research. Wells, Warchal, Ruiz and Chapdelain (in press) indicate that the criteria against which research might be judged in another country could be based on principles that deviate from western or American criteria.

- *Research on your own students*: Because of the power differences between faculty and students, researchers should be careful when undertaking research on their own students, particularly during the semester in which a student is enrolled in the faculty member's course. If a course does have a requirement for research participation, students should be given the opportunity for another option, such as a research paper, without penalty.

- *Research on students in K-12 settings***:** Studies involving minor children require particular care with regard to informed consent. Some procedures, if they are part of the standard teaching and learning environment, do not require separate parental consent. This includes research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special educational instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods. An IRB must review research protocols to determine if and how informed consent will be handled for minors.

Additional Information

Additional information about the ethical issues in service-learning research can be obtained through the following publications:

Chapdelaine, A., Ruiz, A., Warchal, J., & Wells, C. (2005). *Service-learning code of ethics*. Bolton, MA: Anker Publishing Co. http://www.josseybass.com/WileyCDA/WileyTitle/productCd-1882982835.html

Wells, C., Warchal, J., Ruiz, A., & Chapdelain, A. (in press). Ethical issues in research on international service learning. In R. G. Bringle, J. A. Hatcher, & S. G. Jones, (Eds.), *International service learning: Conceptual frameworks and research.* Sterling, VA: Stylus.

Chapter 5
Data Analysis and Interpretation

Once information is gathered in research, it must be organized. The organizational strategy can be structured around observation setting, by theme or variables, or by research questions. The procedures for organizing and reducing data to a form that can be summarized and utilized are different for quantitative and qualitative data. For qualitative information, explicit procedures are used to organize summaries around common themes or categories and then to identify patterns (content analysis). The nature of the qualitative information and the research questions may require a template or rubric, either designed prior to data collection or after examining the data that organizes and summarizes the findings (see Lofland & Lofland, 1995; Miles & Huberman, 1994; Wolcott, 1994). Quantitative information needs to be numerically summarized (e.g., average ratings, frequencies) as well as more extensively analyzed (see Fink, 1995; Fitz-Gibbon & Morris, 1987).

Because there are many ways to conduct the analysis of information, neither quantitative nor qualitative analyses are simple or straightforward, especially to those who do not have experience with statistics or methods of thematic interpretation. Often, it is necessary to try different strategies of data analysis before the most meaningful approach is identified. In addition, different analyses might be appropriate for different purposes or audiences. Consulting an expert in statistics and data analysis is helpful, and hiring an expert to conduct the analysis may be necessary.

As analyses progress, they will need to be put into perspective as to their relevance to the questions or hypotheses that are the focus of the research. The perspective for interpretation might be the theoretical framework, expected results (e.g., hypotheses), a standard or benchmark, a comparison within the set of data (e.g., first-year vs. upper-division students), comparisons over time (e.g., achievement of learning outcomes, changes in attitudes), results from past research (e.g., in the research literature, at your institution), and implications for future programming (e.g., strengths, weaknesses, recommendations). Interpretation of results should be appropriate for the confidence that is inherent in the research design and measurement methods that were selected and should acknowledge limitations, as appropriate. When analyzing and interpreting both quantitative and qualitative data, care should be taken to avoid some of the most *common pitfalls*:

- Assuming that the intervention is the only cause of positive changes documented. Several factors, some of which were unrelated to the intervention, may be responsible for changes in participants or in a community. Isolating specific causes is difficult and the report should at least acknowledge the possibly that other factors may have contributed to change.

- Forgetting that the same methods may give different results when used by different researchers, in different settings, using different procedures, or when different subjects are studied or sampled. For example, two interviewers may ask the same questions but receive different answers because one was friendlier or more patient than the other. As a result, problems or difficulties may be ignored or hidden because people do not report those outcomes.

- Choosing the wrong groups to compare or comparing groups that are different in too many ways. For example, gender, age, race, economic status, and many other factors can all have an impact on outcomes. If comparisons between groups are important, try to compare those with similar characteristics except for the variable being studied.

- Claiming that the results of small-scale research also apply to a wide group or geographic area. For example, it is misleading to claim that participants' responses to a particular intervention in one course apply to the United States as a whole (W. K. Kellogg Foundation, 2006).

Quantitative Analysis

Quantitative analysis of research data is divided into two types, descriptive and inferential statistics. ***Descriptive statistics*** are used to *describe* or summarize the data obtained in the study and to provide an overview of findings. ***Inferential statistics*** are used to make inferences, or draw conclusions that can be extended beyond the immediate data themselves. We will review here the forms of descriptive and inferential statistics most commonly used in service-learning research. Note that this is not intended to be a "how to" discussion, but rather an introduction to the most frequently seen statistics in service-learning research. For more specific information on statistics and their use the reader should reference a statistics text or consult with experienced research colleagues.

Descriptive Statistics

- *Frequency distribution*: A summary of the individual scores or values on a measure or groupings of values, and how frequently that score or value occurred. This can take the form of a table (below), or a figure, such as a histogram, line or bar graph, or pie chart.

Example:	SAT Writing Score	Percent
	200-299	9%
	300-399	15
	400-499	26
	500-599	27

| 600-699 | 13 |
| 700-800 | 10 |

- *Measures of central tendency*: There are three statistics that are used to show the "center" of a distribution. In a normal or bell-shaped distribution these three scores are all equal to each other.
    - *Mean***:** the mathematical average of all scores. The mean is typically used with interval data.
    - *Median***:** the score found at the exact middle of the set of scores. For example, if you have 300 scores and put them in numerical order, the 150th ranked score is the median. The median is particularly useful if there are a few extreme scores that "pull" the mean up or down. The median is appropriate for ordinal data.
    - *Mode***:** the most frequent value in the set of scores; the highest point in the histogram or line graph. Sometimes there is more than one modal value, such as in a *bimodal* distribution. The mode is used for nominal or categorical data.

- *Measures of dispersion*: Dispersion refers to how spread out the scores are in a distribution. There are two common statistics used to show dispersion:
    - *Range*: A simple way to show the "width" of a distribution, the range is highest value minus the lowest value.
    - *Standard deviation*: A descriptive statistic that shows the relationship that the set of scores has to the mean (average) of the distribution. The higher the standard deviation, the bigger the width of the distribution and the more varied the scores are around the mean.

- *Crosstabs***:** A table summarizing combinations of two (or more) characteristics, categories, or scores, and how frequently they occur. In the table below, the two variables being summarized are class status and sex of respondents.

Example: Class	Percent of Males	Percent of Females
Freshmen	49%	46%
Sophomores	25%	30%
Juniors	16%	15%
Seniors	10%	9%

Inferential Statistics

Inferential statistics are used to test hypotheses, make inferences, and draw conclusions that can be extended beyond the immediate data themselves. The most commonly-used inferential statistics in service-learning research are described below. This section also includes a discussion of analysis of pre-test, post-test data, because this is a common measurement strategy in service-learning research.

- *Correlations***:** A correlation demonstrates the nature and degree of association between two naturally occurring variables. The ***correlation coefficient*** is a statistical summary of the nature of the association between two constructs that have been operationalized as variables. The correlation coefficient contains two pieces of information, (a) a number, which summarizes the degree to which the two variables are linearly associated; and (b) a sign, which summarizes the nature of the relationship. The numeric value of a correlation coefficient can range from +1.0 to -1.0. Larger absolute values indicate greater linear association; numbers close to zero indicate no linear relationship. A positive sign indicates that higher values on one variable are associated with higher values on the other variable; a negative sign indicates an inverse relationship between the variables such that higher values on one variable are associated with lower values on the other variable. A correlation coefficient is both a descriptive statistic (i.e., describing the nature of the relationship in a sample) and an inferential statistic (i.e., a sample of the nature of the relationship in a broader population).

- *t-test*: The *t*-test and ***one-way analysis of variance (ANOVA)*** are used to determine if two sets of scores (*t*-test) or two or more sets of scores (ANOVA) are different. One common use of them is to compare the average performance of one group of subjects on a measure before and after a program; either the dependent *t*-test or repeated measure ANOVA can be used to determine if the two sets of scores differ significantly. In this case, the two sets of scores come from the same group of subjects. Another common use is to compare the average scores of one group versus another group, such as the post-test scores of a service-learning class versus the post-test scores of a non-service-learning class. ***ANOVA*** is used when there are two or more groups being compared or when there are more than two independent variables being analyzed.

- *Analysis of covariance (ANCOVA)*: ANCOVA tests whether certain factors (independent variables) have an effect on the dependent variable while statistically removing the effects of other variables (covariates). For example, the researcher might give a pre-test and a post-test to both a service-learning section of a course and to a traditional section that does not include a service component. Because of the possibility of self-selection into the service-learning course, the researcher may wish to control for

prior volunteering. The ANCOVA analysis allows the researcher to control for differences on prior volunteering (i.e., hold statistically constant prior volunteering experience), while examining differences between treatment and non-treatment groups on the dependent variable, thus isolating the effect of the main independent variable on the dependent variable. Another common approach is to use the pre-test as the covariate— i.e., hold the pre-test scores for the two groups constant, and then evaluate whether members of the service-learning group changed more than members of the traditional course section. Because most service-learning research involves non-random assignment of subjects to groups (quasi-experimental), researchers need to use a reliability-corrected ANCOVA model when pre-test scores are available (Trochim, 2006).

- *Multiple regression*: Multiple regression allows the evaluation of the association between a set of independent variables and a dependent variable. Multiple regression can also evaluate the relative importance of the each independent variable to the change in the dependent variable scores. Multiple regression is an improvement over bivariate correlation because multiple regression can examine the association of many predictors (e.g., family background variables, prior volunteering, attitudes, values, moral development) with an outcome variable (e.g., post-graduation civic involvement).

- *Strategies for pre-test/post-test analysis*: One of the most common measurement strategies in service-learning research is to give a measure (e.g., attitude, knowledge) to students at the beginning and the end of the semester to detect change or growth. (See the discussions of Experimental and Quasi-Experimental Designs in Chapter 2.) There are two basic strategies for analysis of pre-test, post-test data. The first strategy is to use a *t*-test to conduct a comparison of differences between ***post-test scores only*** of two groups (e.g., service-learning section versus non-service-learning section). Researchers often use this strategy if they do not have pre-test data available, or if they found no differences on pre-test scores and subsequently choose to ignore the pre-test data in analysis. Unfortunately this strategy suffers from the limitation that it is not possible to conclude that the difference on the post-test is due to the difference in instruction, rather than differences between the groups, general student maturation, or other events external to the course. When pre-test data are available, they should always be included in analyses, even when there are no significant differences between groups on the pre-test scores.

  A second strategy is to analyze the ***raw difference scores*** (post-test minus pre-test scores) for each individual in the groups. This practice is not without controversy (Cronbach & Furby, 1970; Maruish, 1999; Pedhazur & Schmelkin, 1991; Rogosa & Willett, 1983) but is preferable to using post-test scores only, because the researcher is analyzing the change that is occurring for each participant and can make some conclusions, depending on the design of the study, that the changes in scores are due to the educational intervention rather than pre-existing differences in groups. For more

precision the researcher may choose to use blocking, matching, or add a moderator variable such as gender, service site, or some other pre-existing measure (e.g., personality, prior service experience) to the design (Cook & Campbell,1979; Maruish, 1999, Pedhazur & Schmelkin, 1991) and conduct an ANOVA (one dependent variable) or MANOVA (more than one dependent variable) on the difference scores. In this type of analysis, the difference in groups (e.g., intervention) would be one factor (between subjects), "time" would be a factor (within subjects), and the moderator variable would be a factor in the ANOVA or MANOVA analysis. Another option is to conduct a multiple regression or ANCOVA (Edwards, 1994; Pedhazur & Schmelkin, 1991) to statistically remove the effects of moderator or other variables that produce nonequivalence of groups, and to control for pre-test scores. There are other types of scores that have been recommended for pre-, post-test analyses (standardized difference scores, residual change scores), but these are less straightforward, have problems of their own, and are not as appropriate as raw difference scores for research on service-learning.

Qualitative Analysis

Because of the iterative and emergent nature of qualitative research (see Figure 2) it is sometimes difficult to draw a hard line distinguishing data collection from analysis. According to Patton (2002),

> In the course of fieldwork, ideas about directions for analysis will occur. Patterns take shape. Possible themes spring to mind. Hypotheses emerge that inform subsequent fieldwork. While earlier states of fieldwork tend to be generative and emergent, following wherever the data lead, later stages bring closure by moving toward confirmatory data collection—deepening insights into and confirming (or disconfirming) patterns that seem to have appeared. (p. 436)

In contrast to quantitative research, there are no shared ground rules for qualitative analysis, except to represent the data fairly and completely, and to communicate what patterns, themes, and conclusions they reveal. Qualitative analysis involves sifting through large amounts of information, identifying important patterns, and reporting "thick" or rich descriptions of what was found. Patton (2002) identifies several ways to organize and report qualitative data:

- *Storytelling approaches*: chronological, flashback (working backward)
- *Case study approaches*: focus of analysis is on individuals, groups, major events, or settings
- *Analytical framework approaches*: analysis is focused on processes, key issues, topics, concepts, or interview questions

Strategies for Ensuring Validity or Trustworthiness in Qualitative Analysis

Because qualitative research is subjective in nature it is difficult to establish the reliability and validity of the approach and the information produced. Guba (1981) proposed four criteria for judging the "trustworthiness" of qualitative research:

- *Credibility***:** Accomplished by confirming that the results of the research are credible to the participants in the study. The researcher must also establish his or her own credibility by describing any personal or professional information that may have influenced the study (Krefting, 1991; Patton, 2002). This is analogous to *internal validity* in quantitative research.

- *Transferability*: Accomplished in two steps: (a) the investigator must thoroughly describe the context of the research and the assumptions of the study, and (b) the reader or user of the research must decide how well the described study fits another context. This is analogous to *external validity* or *generalizability* in quantitative research.

- *Dependability*: To establish dependability the naturalistic researcher must explain both the stable, consistent elements of research findings, and also the contextual changes that occurred during the study. The researcher must also provide a dense description of the research methodology so that someone else could replicate it, if desired. This is analogous to *reliability* in quantitative research.

- *Confirmability*: The researcher is responsible for describing the research results in such a way that they can be confirmed by others. According to Patton (2002), this can be accomplished in several ways, (a) generating and assessing rival conclusions; (b) finding and analyzing negative cases that contradict prior understandings; (c) triangulating by using multiple methods, sources, analysts, or theories to test for consistency in results; (d) keeping methods and data in context by considering how design constraints may have affected the data available for analysis; and (e) articulating lessons learned and best practices emanating from the research. Confirmability is analogous to *objectivity* in the quantitative approach.

Chapter 6
Dissemination of Research Results

Dissemination Avenues

In order for others to know about the outcomes of a research project, the researcher must summarize the study for dissemination. There are many avenues to disseminate research results. Some of these include:

- Reports to funders
- Reports to community partners
- Presentations to students, institution, and community partners
- Presentations at professional conferences
- Research briefs
- Website publication (e.g., department website)
- Peer-reviewed journal article
- Chapter in an edited book
- Articles in newsletters
- Dissertation or thesis

Qualities of Good Research Articles (Quantitative)

The value of assessment, program evaluation, and research lies in generating information in a systematic way to inform others about the impact of educational programs on various constituencies. Therefore, reports on the results of research should be tailored to particular audiences and purposes. Researchers may, therefore, prepare more than one report of the findings and conclusions (e.g., executive summary, detailed research report, short paragraphs for a newsletter, manuscript for publication in a journal, Power Point slides for a presentation). Each product should contain the necessary details that the intended audience needs, with graphic and tabular representations of the findings, as appropriate. The conclusions and recommendations should distinguish between conclusions based on robust data and those that are more speculative. Promises of confidentiality should always be honored.

Empirical Research Reports: Journals

Reports of research that will appear in academic journals need to follow the journal's guidelines. These reports will typically have an hour-glass shape (on its side): they start rather general, becoming more specific in detailing the research question (e.g., hypothesis) addressed in the research; they are very specific with regard to methods and findings; then they become

progressively more general as the theoretical and practical implications of the research are presented and discussed. The ***introduction*** and ***discussion*** sections should be easily read by a general audience, whereas the ***methods*** and ***results*** sections are more technical. The Introduction section is sometimes referred to as a literature review; however, good Introductions are not exhaustive reviews of far-ranging literature, but focus on presenting a rationale for each research question or hypothesis that is addressed in the research. The methods section should provide sufficient detail so that the study could be replicated by someone else. This section should clearly state what the design and procedures of the study were, the manner in which sampling occurred, whether or not random assignment was used, and the psychometric properties of measures used (Wilkinson, 1999). The results section should provide a logical and coherent presentation of the rationale for data analysis and those results that are pertinent to evaluating the research questions. All data analyses that were conducted do not need to be reported; researchers should select those that provide a coherent evaluation of the research questions.

The discussion section should start with a one paragraph summary of the purpose and procedures of the research. Subsequent sections may be organized as going from specific issues (e.g., the research project and particular findings) to more general issues (e.g., implications of the research). The discussion *should not* merely repeat the results. The discussion should contain the (a) conceptual and theoretical implications of the results, (b) the practical implications of the results, and (c) the connections of the results to past research. This is the case for both significant findings *and non-significant* findings. (See the APA Publication Manual (2010) for contents of the discussion.) Null or non-significant findings can be presented in terms of (a) the theory, and thus the hypothesis, being wrong; (b) there were measurement problems that prevented a reasonable test of the hypothesis (e.g., a scale had an unacceptable coefficient alpha); (c) the design of the study was flawed or inadequate; (d) the execution of the research deviated from the protocol; or (e) a sampling issue contributed to the null results (sample size, nature of sample). Writing the discussion may require that new literature be cited to interpret unexpected findings. The discussion should be very cautious about making *causal* statements, although it may explore these issues. The discussion may contain a discussion of the limitations of the research, but this should not be a major part of the discussion.

Publication and Presentation Outlets for Research on Service-Learning

The following are potential outlets for scholars hoping to disseminate results from their service-learning research:

Michigan Journal of Community Service Learning **(MJCSL)** (www.umich.edu/~mjcsl/)
Since 1994, the *Michigan Journal of Community Service Learning* (MJCSL) has been the premiere national, peer-reviewed journal publishing articles written by faculty and service-learning educators on research, theory, pedagogy, and other issues related to academic (curriculum-based) service-learning in higher education.

Advances in Service-Learning Research (www.infoagepub.com/products/series/s0007.php)
This book series gathers current research on service-learning in K-12 education, teacher education, and higher education. Along with chapters highlighting the findings of service-learning research studies, the book series includes thought pieces that identify theoretical groundings of service-learning and present methodological approaches for studying service-learning (including teacher action research). Contributions to these volumes are limited to research presented at the annual meetings of the International Association for Research on Service-Learning and Civic Engagement.

Journal of Community Engagement and Scholarship (www.jces.ua.edu)
The *Journal of Community Engagement and Scholarship* addresses critical problems identified through community-based participatory research, a rapidly developing approach to cutting-edge scholarship in which students and community partners play important roles.

Partnerships: A Journal of Service-Learning & Civic Engagement
(www.partnershipsjournal.org)
Partnerships recognizes that successful engaged learning depends on effective partnerships between students, faculty, community agencies, administrators, disciplines, and more. The articles in this peer-reviewed journal focus on how theories and practices can inform and improve such partnerships, connections, and collaborations. Studies co-authored by faculty, students, and/or community partners; or examining practices across disciplines or campuses; or exploring international networks are all encouraged.

Research in Service Learning: Publishing Opportunities Resource List
(www.civicyouth.org/PopUps/SR_Homana.pdf)
This resource list is published by the Center for Information & Research on Civic Learning & Engagement (CIRCLE). CIRCLE includes a list of 93 potential research publishing outlets and annotated information on each of the publications.

List of Publishing Outlets for Service-Learning and Community-Based Research
(www.compact.org/category/resources/service-learning-resources/publishing-outlets-for-service-learning-and-community-based-research/)
Compiled by Campus Compact, this resources list includes 43 publishing outlets for service-learning and community-based research.

Publishing and Presenting in Service-Learning
(servicelearning.gov/instant_info/fact_sheets/he_facts/publishing_sl)
Compiled by the National Service-Learning Clearinghouse, this source provides a number of resources for publishing and presenting service-learning research results.

ERIC: Information for Individual Contributors
(www.eric.ed.gov/ERICWebPortal/resources/html/submission/content_providers_c_overview.html)
The Education Resources Information Center (ERIC) is an online digital library of education research and information. Individuals may submit high-quality, education-related materials, including copyrighted materials, for inclusion in the ERIC database. Types of materials appropriate for individual submission include research reports, conference papers and presentations, and dissertations and theses. ERIC does not accept lesson plans, blogs, or individual Web pages.

Publication Outlets for the Scholarship of Teaching and Learning

The following are outlets for publishing research on teaching and learning, including, but not limited to, research on service-learning.

Journal of the Scholarship of Teaching and Learning (www.iupui.edu/~josotl)
The *Journal of the Scholarship of Teaching and Learning* (JoSoTL) is a forum for the dissemination of the Scholarship of Teaching and Learning in higher education for the community of teacher-scholars. The peer-reviewed journal promotes SoTL investigations that are theory-based and supported by evidence. JoSoTL's objective is to publish articles that promote effective practices in teaching and learning and add to the knowledge base.

Journal of Community Engagement and Higher Education (www.indstate.edu/jcehe)
The Journal of Community Engagement and Higher Education is an on-line, refereed journal examining perspectives, research, and practices of community engagement and community-based learning in higher education.

Journal of Higher Education Outreach and Engagement
(openjournals.libs.uga.edu/index.php/jheoe/)
The *Journal of Higher Education Outreach and Engagement* is a peer-reviewed publication whose editorial goals are to serve as a forum to promote the continuing dialogue about the service and outreach mission of the university; and to foster understanding of how the service and outreach mission relates to the university's teaching and research missions as well as the needs of the sponsoring society.

International Journal for the Scholarship of Teaching & Learning
(academics.georgiasouthern.edu/ijsotl)
International Journal for the Scholarship of Teaching & Learning (IJ-SoTL) is an open, peer-reviewed, international electronic journal published twice a year by the Center for Excellence in Teaching at Georgia Southern University. The journal strives to be an international vehicle for

articles, essays, and discussions about the scholarship of teaching and learning (SoTL) and its applications in higher/tertiary education today. All submissions undergo a double-blind peer-review process.

Journal of Applied Learning in Higher Education
(www.missouriwestern.edu/AppliedLearning/journal.asp)
The *Journal of Applied Learning in Higher Education* (JALHE) serves the international community of scholars engaged in applied learning at institutions of higher education. Topics may focus on assessment or evaluation of the quality of applied learning, the development or administration of applied learning programs, representing and valuing applied learning for promotion and tenure processes, the relationship between applied learning practice and student learning outcomes, or related topics.

References

American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Retrieved from http://www.apa.org/ethics/code/index.aspx#general

American Psychological Association. (2010). *Publication manual of the American Psychological Association (6ᵗʰ Ed.)*. Washington, DC: APA.

Annette, J. (2003). Service-learning internationally: Developing a global civil society. In S. Billig & J. Eyler (Eds.), *Deconstructing service-learning: Research exploring context, participation, and impacts* (pp. 241-249). Greenwich, CN: Information Age.

Ash, S. L., & Clayton, P. H. (2004). The articulated learning: An approach to guided reflection and assessment. *Innovative Higher Education*, *29*(2),137-154.

Ash, S. L., Clayton, P. H., & Atkinson, M. P. (2005). Integrating reflection and assessment to capture and improve student learning. *Michigan Journal of Community Service Learning, 11*(2), 49-60.

Astin, A. W., & Sax, L. J. (1998). How undergraduates are affected by service participation. *Journal of College Student Development, 39*, 251-263.

Astin, A. W., Sax, L. J., & Avalos, J. (1999). The long-term effects of volunteerism during the undergraduate years. *The Review of Higher Education*, *21*(2), 187-202.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Battistoni, R. M. (2002). *Civic engagement across the curriculum: A resource book for service-learning faculty in all disciplines*. Providence, RI: Campus Compact.

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics*. New York, NY: Oxford University Press.

Brandt, A. M. (1978). Racism and research: The case of the Tuskegee Syphilis Study. *Hastings Center Report, 8*(6), 21-29.

Bringle, R. G. (2003). Enhancing theory-based research on service-learning. In S. H. Billig & J. Eyler (Eds.), *Deconstructing service-learning: Research exploring context, participation, and impacts* (pp. 3-21). Greenwich, CN: Information Age Publishing.

Bringle, R. G., & Hatcher, J. A. (1995). A service-learning curriculum for faculty. *Michigan Journal of Community Service Learning*, *2*, 112-122.

Bringle, R. G., & Hatcher, J. A. (1996). Implementing service learning in higher education. *Journal of Higher Education*, *67*, 221-239.

Bringle, R. G., & Hatcher, J. A. (1999). Reflection in service learning: Making meaning of experience. *Educational Horizons, 77*(4), 179-185.

Bringle, R. G., & Hatcher, J. A. (2000, Fall). Meaningful measurement of theory-based service-learning outcomes: Making the case with quantitative research. *Michigan Journal of Community Service Learning*, 68-75.

Bringle, R. G., & Hatcher, J. A. (2009, Fall). Innovative practices in service-learning and curricular engagement. *New directions for higher education. No. 147,* 37-46.

Bringle, R. G., Phillips, M. A., & Hudson, M. (2004). *The measure of service learning: Research scales to assess student experiences.* Washington, DC: American Psychological Association.

Bringle, R. G., & Steinberg, K. (in press). Educating for informed community involvement. *American Journal of Community Psychology*.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago, IL: Rand McNally.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Skokie, IL: Rand McNally.

Campus Compact. (2006). *Campus Compact annual membership survey*. Retrieved from http://www.compact.org/about/statistics/2006/

Chapdelaine, A., Ruiz, A., Warchal, J., & Wells, C. (2005). *Service-learning code of ethics*. Bolton, MA: Anker Publishing.

Code of Federal Regulations. (2005). Title 45-Public welfare. Department of Health and Human Services. Part 46. *Protection of human subjects*. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm#46

Cook, T., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston, MA: Houghton Mifflin.

Cozby, P.C. (2009). *Methods in behavioral research, 10th Ed.* Boston: McGraw-Hill.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—Or should we? *Psychological Bulletin, 74*, 68-80.

Denzin, N., & Lincoln, Y. S. (2005). (Eds.). *Handbook of qualitative research (3rd Ed.).* Thousand Oaks, CA; Sage.

Deutsch, M., & Krauss, R. M. (1965). *Theories in social psychology*. New York, NY: Basic Books.

Dionne, E. J., Jr., & Drogosz, K. M. (2003, Winter). United we serve: The debate over national service. *Campus Compact Reader*, 23-26.

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal of Management, 20(3),* 683-690.

Enos, S., & Troppe, M. (1996). Service-learning in the curriculum. In B. Jacoby (Ed.), *Service-learning in higher education* (pp. 156-181). San Francisco, CA: Jossey-Bass Publishers.

Eyler, J. (2002). Stretching to meet the challenge: Improving the quality of research to improve the quality of service-learning. In S. H. Billig & A. Furco (Eds.), *Service-learning: Through a multidisciplinary lens. Advances in service-learning research series* (pp. 3-14). Greenwich, CT: Information Age.

Eyler, J., Giles, D. E., & Schmiede, A. (1996). *A practitioner's guide to reflection in service-learning: Student voices and reflections.* Nashville, TN: Vanderbilt University.

Eyler, J. & Giles, D. E. (1999). *Where's the learning in service-learning?* San Francisco, CA: Jossey-Bass.

Fink, A. (1995). *How to ask survey questions*. Thousand Oaks, CA: Sage.

Fitz-Gibbon, C. T., & Morris, L. L. (1987). *How to analyze data*. New York, NY: Sage.

Furco, A. (1996). Service-learning: A balanced approach to experiential education. In Corporation for National Service, *Expanding boundaries: Serving and learning* (pp. 2-6). Columbia, MD: The Cooperative Education Association.

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction.* Los Angeles: Sage.

Gelmon, S. B., Furco, A., Holland, B. A., & Bringle, R. G. (2005, November 13). *Beyond anecdote: Challenges in bringing rigor to service-learning research.* Panel presented at the Fifth International Service-Learning Research Conference, East Lansing, MI.

Gelmon, S. B., Holland, B. A., Driscoll, A., Spring, A, & Kerrigan, S. (2001). *Assessing service-learning and civic engagement: Principles and techniques.* Providence, RI: Campus Compact.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Technology Research and Development, 29*(2), 75-91.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research (3rd Ed.).* Thousand Oaks, CA: Sage.

Hatcher, J. A., & Bringle, R. G. (1997). Reflection: Bridging the gap between service and learning. *Journal of College Teaching, 45*, 153-158.

Howard, J. (1993). Community service in the curriculum. In J. Howard (Ed.), *Praxis I: A faculty casebook on community service learning* (pp. 3-12). Ann Arbor, MI: OCSL Press.

International Association on Research on Service-Learning and Community Engagement. (2009). *About us: Mission*. Retrieved from http://www.researchslce.org/_Files/Public_Site/About_Us_Files/aboutus.html

Keily, R., & Hartman, E. (in press). Qualitative research methodology and international service learning: Concepts, characteristics, methods, approaches, and best practices. In R. G. Bringle, J. A. Hatcher, & S. G. Jones (Eds.), *International service learning: Conceptual frameworks and research.* Sterling, VA: Stylus.

Kerlinger, F. N. (1986). *Foundations of behavioral research (3rd Ed.).* New York: Holt, Rinehart & Winston.

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy, 45*(3), 214-222.

Levine, P. (2003, November). *Service-learning research and the movement for youth civic engagement*. Keynote address at the Third Annual International Service-Learning Research Conference, Salt Lake City, UT.

Lofland, J., & Lofland, L. (1995). *Analyzing social settings: A guide to qualitative observation and analysis*. Belmont, CA: Wadsworth.

Markus, G. B., Howard, J. P. F., & King, D. C. (1993). Integrating community service and classroom instruction enhances learning: Results from an experiment. *Educational Evaluation and Policy Analysis*, *15*, 410-419.

Maruish, M. E. (1999). *The use of psychological testing for treatment planning and outcomes assessment.* Mahwah, NJ: Lawrence Erlbaum.

McGuire, W. J. (1980). The development of theory in social psychology. In S. Duck & R. Gilmour (Eds.), *The development of social psychology* (pp. 53-80). London, UK: Academic Press.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research*. Retrieved from http://ohsr.od.nih.gov/guidelines/belmont.html

Nunnally, J. C. (1967). *Psychometric theory*. New York, NY: McGraw-Hill.

Nuremberg Code. (1949). In *Trials of war criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol.2* (pp. 181-182). Washington, DC: U.S. Government Printing Office. Retrieved from http://ohsr.od.nih.gov/guidelines/nuremberg.html

Osborne, R. E., Hammerich, S., & Hensley, C. (1998). Student effects of service-learning: Tracking change across a semester. *Michigan Journal of Community Service Learning, 5,* 5-13.

Patton, M. Q. (2002). *Qualitative research & evaluation methods (3 rd Ed.).* Thousand Oaks, CA: Sage Publications.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum.

Perold, H. (1998). *Community service in higher education: Final report*. Johannesburg, South Africa: Joint Education Trust.

Pimple, K. D. (2006, November). *Protection of human subjects in non-biomedical research: A tutorial*. Retrieved from http://www.indiana.edu/~rcr/files/hspt-nbm.pdf

Rogoza, D. R., & Willett, B. (1983). Demonstating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement, 20*, 335-343.

Rothman, D. J. (1982). Were Tuskegee & Willowbrook 'studies in nature'? *Hastings Center Report, 12*(2), 5-7.

Schneider, W. H. (n.d.). *The establishment of institutional review boards in the U.S. background history*. Retrieved from http://www.iupui.edu/~histwhs/G504.dir/irbhist.html

Schwandt, T. (2001). *Dictionary of qualitative inquiry (2nd Ed.).* Thousand Oaks, CA: Sage.

Shaw, M. E., & Costanzo, P. R. (1982) *Theory in social psychology*. New York, NY: McGraw-Hill.

Steinke, P., & Buresh, S. (2002). Cognitive outcomes of service-learning: Reviewing the past

and glimpsing the future. *Michigan Journal of Community Service Learning, 8*(2), 5-14.

Trochim, W. M. (2006). *The research methods knowledge base (2ndEd.).* Retrieved from
http://www.socialresearchmethods.net/kb/

Wells, C., Warchal, J., Ruiz, A., & Chapdelain, A. (in press). Ethical issues in research on
international service learning. In R. G. Bringle, J. A. Hatcher, & S. G. Jones, (Eds.),
International service learning: Conceptual frameworks and research. Sterling, VA:
Stylus.

Wilkinson, L. (1999, August). Statistical methods in psychology journals: Guidelines and
explanations. *American Psychologist, 54*(8), 594-604.

W. K. Kellogg Foundation. (2006). *2005 Annual report: From ideas to action.* Retrieved from
http://www.wkkf.org/knowledge-center/resources/2006/12/WK-Kellogg-Foundation-
2005-Annual-Report-From-Ideas-To-Action.aspx

Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation.*
Thousand Oaks, CA: Sage.

Appendix 1:
Service-Learning Research Resources on the Web

Campus Compact National Office (www.compact.org/)
Campus Compact is a national coalition of more than 1,100 college and university presidents dedicated to promoting community service, civic engagement, and service-learning in higher education. The web site provides resources and publications related to service-learning, assessment, campus-community partnerships, civic engagement, community colleges, faculty, college administration, program models, promotion and tenure, student leadership, and a variety of other topics.

Center for Information and Research on Civic Learning and Engagement
(www.civicyouth.org/)
The Center for Information and Research on Civic Learning and Engagement (CIRCLE) conducts research on the civic and political engagement of Americans between the ages of 15 and 25. CIRCLE is based at the Jonathan M. Tisch College of Citizenship and Public Service at Tufts University. Resources for researchers include data sets, fact sheets, a working paper series, special reports and books, research newsletter, and links to other research sites.

Compendium of Assessment and Research Tools (cart.rmcdenver.com/)
The Compendium of Assessment and Research Tools (CART) is a database that provides information on instruments that measure attributes associated with youth development programs. CART includes descriptions of research instruments, tools, rubrics, and guides and is intended to assist those who have an interest in studying the effectiveness of service-learning, safe and drug-free schools and communities, and other school-based youth development activities.

Community-Campus Partnerships for Health (depts.washington.edu/ccph/index.html)
Community-Campus Partnerships for Health (CCPH) is a nonprofit organization that promotes health (broadly defined) through partnerships between communities and higher educational institutions. Online resources include publications and materials related to service-learning, community-based participatory research, community-engaged scholarship, and funding opportunities (updated every two weeks).

Corporation for National and Community Service (www.nationalservice.gov/)
The Corporation for National and Community Service (CNCS) was formed to engage Americans of all ages and backgrounds in service to meet community needs. Corporation's three major programs are Senior Corps, AmeriCorps, and Learn and Serve America. The Corporation publishes studies on volunteering, civic engagement, service-learning, and the effectiveness and outcomes of its programs.

International Association for Research on Service-Learning and Community Engagement (www.researchslce.org/)
The International Association for Research on Service-Learning and Community Engagement (IARSLCE) is an international, non-profit organization devoted to promoting research and discussion about service-learning and community engagement. IARSLCE holds an annual research conference and provides networking opportunities for people interested in service-learning research. It publishes *Advances in Service-Learning Research,* a series of volumes developed from the annual research conference.

IUPUI Center for Service and Learning Research Collaborative (csl.iupui.edu/cslrc.asp)
The Center for Service and Learning Research Collaborative (CSLRC) provides resources, convenes scholars, conducts and publishes research, and disseminates information related to research on service-learning. The CSLRC established a partnership with Learn and Serve America's National Service-Learning Clearinghouse (NSLC) in order to create a web portal on service-learning research. The CSLRC contracted with Stylus Publishing to create the new *IUPUI Series on Service Learning Research*. In addition, the CSLRC hosts the IUPUI Research Academy, an annual workshop on service-learning research.

Learn and Serve America's National Service-Learning Clearinghouse (servicelearning.gov)
The National Service-Learning Clearinghouse (NSLC) supports the service-learning community in kindergarten through higher education settings, community-based initiatives, as well as all others interested in strengthening schools and communities using service-learning techniques and methodologies. The Clearinghouse maintains a website to support service-learning programs, practitioners, and researchers. The website include a set of research tools (servicelearning.gov/service-learning-research-tools/what-research-introduction), designed by the IUPUI Center for Service and Learning, to provide information and support to scholars conducting research on service-learning. In addition, the Clearinghouse operates national email discussion lists to encourage discussion and exchange of ideas. The Clearinghouse also maintains a library collection of print and online resources, and offers reference assistance related to program startup, academic research, assessment and evaluation, and service-learning topics.

RMC Research (www.rmcdenver.com/)
RMC Research is a national leader in program research and evaluation, professional development, consultation, and product development. From small studies to multi-year quantitative research projects, it supports national, state, and local clients who serve schools, families, and communities. Services from their six offices can be customized to meet specific client needs.

Appendix 2:
Online Resources on Research Methodology and Statistics

There are many resources available online and in print on the topics of research design, methodology, and statistics. A few online sources are listed below:

General research methods:
- Research Methods Knowledge Base (www.socialresearchmethods.net/kb/)
- Research Methods Knowledge Base: Design (www.socialresearchmethods.net/kb/design.php)
- Research Methods Tutorials (www.socialresearchmethods.net/tutorial/tutorial.htm)
- Research Methods Knowledge Base: Survey Research (www.socialresearchmethods.net/kb/survey.php)

Statistics:
- Social Psychology Network: Research Methods and Statistics Links by Subtopic: Statistics and Data Analysis (www.socialpsychology.org/methods.htm#generalstats)
- Introductory Statistics: Concepts, Models, and Applications (www.psychstat.missouristate.edu/sbk00.htm)
- Interactive Statistical Calculation Pages (www.statpages.org/)

Statistics Software:
- SPSS (www.spss.com/software/statistics/)
- SAS (www.sas.com/technologies/analytics/statistics/stat/index.html)
- Stata: (www.stata.com/)
- BDMP (www.statistical-solutions-software.com/products-page/bmdp-statistical-software/)

Data Analysis in Excel:
- Analysis ToolPak (office.microsoft.com/en-us/excel-help/perform-statistical-and-engineering-analysis-with-the-analysis-toolpak-HP010090842.aspx?pid=CH100648511033)

Qualitative Methods:
- Qualitative Research Resources on the Internet (www.nova.edu/ssss/QR/qualres.html)
- Research Methods Knowledge Base: Qualitative Measures (www.socialresearchmethods.net/kb/qual.php)

Qualitative Software:

- NVivo (www.qsrinternational.com/products_nvivo.aspx)
- American Evaluation Association: Qualitative Software (www.eval.org/Resources/QDA.htm)
- CAQDAS Comparison (www.restore.ac.uk/lboro/research/software/caqdas_comparison.php)
- Choosing a CAQDAS Package (caqdas.soc.surrey.ac.uk/PDF/2009ChoosingaCAQDASPackage.pdf)

Appendix 3:
Potential Funders for Service-Learning Research Projects

- AAC&U: Bringing Theory to Practice Grants
  (www.aacu.org/bringing_theory/index.cfm)
- CIRCLE (www.civicyouth.org/)
- Institute for Research on Unlimited Love
  (www.unlimitedloveinstitute.org/grant/index.html)
- Lumina Foundation
  (www.luminafoundation.org/grants/information_for_grant_seekers/index.html)
- NASPA Foundation (www.naspa.org/fdn/grants.cfm)
- National Science Foundation: Course, Curriculum and Laboratory Improvement
  (www.nsf.gov/funding/)
- Spencer Foundation (www.spencer.org/content.cfm/research)
- John Templeton Foundation (www.templeton.org/)
- U.S. Department of Education: National Center for Education Research (ies.ed.gov/ncer/)