Chapter 8 Evaluation Research Design: Options

There are two general classifications of evaluation research designs: quantitative and qualitative. Quantitative designs tend to produce numerical output; whereas qualitative designs express their findings as narrative. Increasingly, evaluators are using a blended approach, drawing designs from either "toolbox," and/or specialized designs to answer evaluation research questions.

Essentially, quantitative research is a collection methods (experimental, causal-comparative, correlation, and survey research) used to inquire into a problem, issue, question, theory, etc. of interest to a researcher or research team. A question or theory, composed of variables, is measured in a systematic way and data are analyzed with statistical procedures. Chapters 3 through 8, Part 1 focus on quantitative research.

Qualitative research is a system of inquiry which seeks to build a holistic, largely narrative, description to inform the researcher's understanding of a social or cultural phenomenon. Qualitative research takes place in natural settings employing a combination of observations, interviews, and document reviews. Chapter 8, Part II briefly examines qualitative research.

Following is a discussion of (1) specific quantitative (pre-experimental, quasi-experimental, and true-experimental) research designs, including their internal and external design threats, and statistical analysis procedures; (2) survey research; (3) six qualitative research designs; and (4) two specialized designs.

I. Pre-experimental, Quasi-experimental, True-Experimental & Survey Designs

There are three traditional classifications of quantitative evaluation research designs (or data collection strategies): pre-experimental, quasi-experimental & true-experimental. Pre-experimental designs are considered less rigorous than the true-experimental designs, which many consider the "Gold Standard" and referred to as RCTs or Randomized Control Trials. Pre-experimental designs are the quantitative research designs most closely associated with action or applied research. Findings or results from pre-experimental designs aren't generalizable to other samples or groups, due to a lack of randomization in subject selection or assignment. There is a fairly standard set of symbols for graphing research designs; the symbols relevant to our discussion are found in Table 8.1.

Table 8.1
Quantitative Evaluation Research Design Symbols

Symbol	Meaning
R	Random assignment or selection sample members; used in True Experimental Designs
X	Treatment or independent variable; used in all quantitative designs
O	Observation or a data collection instance, used in all quantitative designs. A subscript indicates observation number. A dependent viable is measured. Data collection may involve one or multiple data collection tools.
C	Control group in experimental designs
C	Comparison group in pre-experimental or quasi-experimental designs

If using the Kirkpatrick or another evaluation framework, it may be necessary to use different study designs for each level. For example, Kirkpatrick Level 1, may be measured using a survey instrument employing the One-Shot Case Study Design (Figure 8.1). Level 2 might be measured with a traditional test (knowledge) and a direct performance assessment (skill) using the One Group Pretest-Posttest Design (Figure 8.2). Level 3 might be assessed, i.e., measured, using checklists or a direct performance assessment using the Static Group Comparison Design (Figure 8.3). Level 4 (unit or organization-wide) effects are usually measured using financial or other performance metrics.

For those interested in more detailed examinations of research design, consult Best & Kahn (2006); Bordens and Abott (2010); Christensen, Johnson, and Turner (2010); Creswell (2013); McMillan and Schumacher (2010); or Cooper and Schindler (2013). Remember, the bibliography presented in Appendix 6.1.

A. Pre-experimental Designs

1. Figure 8.1 One-Shot Case Study Design
   a. Diagram

      Group X O_1

   b. The "group" in this design tends to be purposeful, i.e., a convenience or naturally occurring (as in a classroom or corporate training class) group. Purposeful sampling strategies are most often used with this design; so, if this design is utilized, focus primarily on threats to internal design validity. Do not confuse this design with the qualitative case study presented later.
   c. Probable Operating Threats to Internal Validity. History, mortality, and maturation are likely to operate if the time between application of the independent variable and measurement is extended. Selection is likely to be a problem as the group is not randomly selected, which limits generalizability beyond the sample tested; also, the "sample" may not represent the population from which it was "drawn."
   d. Non-Applicable Internal Validity Threats. Statistical regression, testing, instrumentation, and selection/maturation interaction are unlikely threats as the measure is administered once. The time between treatment and measurement is typically not long in this design; however, if the duration is extended, then selection/maturation interaction may pose a threat.
   e. Possible Operating Threats to External Validity. The reactive or interactive effect of testing (i.e. pre-testing experience), the interactive effects of the treatment with selection, instrumentation, history, or maturation, subject reactive effects to the experimental arrangements, multiple treatment interference (common in longitudinal designs or designs multiple measurements), and experimenter effects likely operate when samples or groupings are not randomly selected. Hence, generalizability beyond the group is not recommended.

f. Statistical Analysis. Descriptive statistics should be routinely applied. Since one observation is taken the z-test or one-sample t-test procedure is appropriate.
g. Comments
   (1) A very weak design as virtually none of the internal validity threats is controlled. This design is employed when one is attempting to describe or test the relationship between an antecedent and a consequent.
   (2) This means that there are several alternative explanations for the effect of the independent variable (or treatment) on the dependent variable. Thus, we are unable to assert with any real certainty that the independent variable exerted an influence on the dependent variable.
   (3) It is unwise to generalize beyond the sample selected given the operating internal validity threats and possible operating threats to external validity.

2. Figure 8.2 One-Group Pretest-Posttest Design
   a. Diagram

      Group O_1 X O_2

   b. The "group" in this design tends to be purposeful, i.e., a convenience or naturally occurring (as in a classroom or corporate training class) group. Purposeful sampling strategies are most often used with this design; so, if this design is utilized, focus primarily on threats to internal design validity.
   c. Probable Operating Threats to Internal Validity. History, mortality, and maturation are likely to operate if the time between application of the independent variable and measurement is extended. Selection is likely to be a problem as the group is not randomly selected, which limits generalizability beyond the sample tested; also, the "sample" may not represent the population from which is was "drawn." Regression maybe a concern as high and/or low outlier scores tend to migrate towards the group mean, when multiple measurements are taken.
   d. Non-Applicable Internal Validity Threats. If the group is purposeful and the time between the first and second measurement is reasonably short, then selection and mortality most likely don't operate. However, if the "group" is not representative of its population and generalization is intended and as the time between measurements increases, selection and mortality are increasingly likely to operate.
   e. Possible Operating Threats to External Validity. The reactive or interactive effect of testing (i.e. pre-testing experience), the interactive effects of the treatment with selection, instrumentation, history, or maturation, subject reactive effects to the experimental arrangements, multiple treatment interference (common in longitudinal designs or designs multiple measurements), and experimenter effects likely operate when samples or groupings are not randomly selected. Hence, generalizability beyond the group is not recommended.

f. Statistical Analysis. Descriptive statistics should be routinely applied. Since two observations are taken, the two sample z-test or dependent samples t-test procedures are appropriate.

g. Comments
   (1) The comments above apply here as well. The difference is that this design is typically employed to assess the influence of an independent variable on a dependent variable. While there is a measure of change generated there are many possible rival explanations.
   (2) This design is very common and quite deceptive as it gives the illusion of design rigor because of the pretest.
   (3) It is unwise to generalize beyond the sample selected given the operating internal validity threats and possible operating threats to external validity.

3. Figure 8.3 Static Group Comparison Design
   a. Diagram

      Group 1 X O_1

      Group 2 O_2

   b. The "group" in this design tends to be purposeful, i.e., a convenience or naturally occurring (as in a classroom or corporate training class) group. Purposeful sampling strategies are most often used with this design; so, if this design is utilized, focus primarily on threats to internal design validity.
   c. Probable Operating Threats to Internal Validity. Selection, maturation, mortality, & selection/maturation interaction are likely to operate as outlined previously. With the addition of a second group, resentful demoralization of the control group, and diffusion of treatment may operate. Group two, the comparison group, does not receive the treatment which may cause resentment and affect performance on the post-test. If study subjects are "close" then treatment effects may cross over to the group 2.
   d. Probable Controlled Internal Validity Threats. Since the measure is administered once, testing, instrumentation, & statistical regression don't operate. History may exert an influence on the dependent variable if there are significant related events, but the time horizon for this design is typically short.
   e. Possible Operating Threats to External Validity. The interactive effect of the treatment with selection, history, or maturation; multiple treatment interference, experimenter effects; and/or reactive or interactive effect of testing may operate as subjects where not randomly selected or assigned.
   f. Statistical Analysis. Descriptive statistics should be routinely applied. Since two observations are taken, the two sample z-test or independent samples t-test procedures are appropriate.

g. Comments
   (1) This is the most reliable of the pre-experimental designs. The purpose of this design is to assess the impact of an independent variable or treatment on one group, and the failure to apply the IV or treatment on the second.
   (2) As with the two above designs, there is no assurance that either group is equal or even representative of its larger population or that they were equal on the dependent variable prior to application of the independent variable or treatment.
   (3) It is unwise to generalize beyond the sample selected given the operating internal validity threats and possible operating threats to external validity.

B. Quasi-Experimental Designs

Quasi-experimental designs are more rigorous than pre-experimental designs because a comparison group is added. The comparison group let's a researcher assert with more certainty that any change in the treatment group's dependent variable is due to their exposure to the independent variable. The comparison group was not exposed to the independent variable. Generalizing study findings or results beyond the groups studied isn't usually done as subjects were not randomly selected for the study sample or assigned to either the treatment or comparison groups.

1. Figure 8.4 Non-equivalent Control Group Design
   a. Diagram

Group 1	O_1	X	O_3
Group 2	O_2		O_4

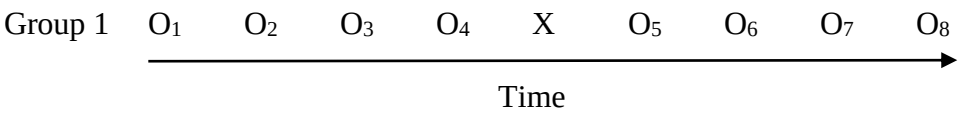
   b. The group in this design tends to be purposeful, i.e., a convenience or naturally occurring (as in a classroom or corporate training class) group. Purposeful sampling strategies are most often used with this design; so, if this design is utilized, focus primarily on threats to internal design validity.
   c. Probable Operating Threats to Internal Validity. Resentful demoralization of the control group and diffusion of treatment, and selection/maturation interaction may operate given the presence of a comparison group and lack of random selection and/or assignment, respectively. Regression may be a concern, given multiple measurements.
   d. Probable Controlled Internal Validity Threats. History, maturation, testing, instrumentation, regression, selection, and mortality are most likely controlled given the presence of a comparison group as in theory what affects the Group 1 will likely affect Group 2 in the same or similar manner.
   e. Possible Operating Threats to External Validity. The interactive effect of the treatment with selection, history, or maturation; multiple treatment

interference, experimenter effects; and/or reactive or interactive effect of testing may operate as subjects were not randomly selected or assigned.

f. Statistical Analysis. Descriptive statistics should be routinely applied. Since two observations are taken the two sample z-test or independent samples t-test procedures are appropriate. If two or more groups are employed, the Analysis of Variance (ANOVA) or Analysis of Covariance (ANCOVA) procedures may be applied. The ANCOVA is particularly useful as it can statistically "control" for specific extraneous or moderating variables (Stevens, 1999, pp.307-308).

g. Comments
   (1) Used when random selection and/or assignment is not possible. It's best to use "in-tact" or naturally occurring groups. Avoid self-selected samples with this design. Of the quasi-experimental designs, this is the strongest. A sound strategy for strengthening this design is to compare or test pre-test results. If no statistically significant results are found, then the argument that the two comparisons groups are equivalent on the dependent variable is stronger.
   (2) The term "comparison group" is used to denote the lack of randomization.
   (3) If the design is used, try to match the two groups on as many relevant variables as possible. This will reduce some of the possible initial differences between the groups.

2. Figure 8.5a Time Series Design
   a. Diagram



   b. The group in this design tends to be purposeful, i.e., a convenience or naturally occurring (as in a classroom or corporate training class) group. Purposeful sampling strategies are most often used with this design; so, if this design is utilized, focus primarily on threats to internal design validity.
   c. Probable Operating Threats to Internal Validity. History as such studies are conducted over an extended period of time, and instrumentation as any design defect will likely exert an influence over multiple measurements may operate.
   d. Probable Controlled Internal Validity Threats. History, testing, regression, selection, mortality, and selection/maturation interaction likely don't operate. Since most longitudinal studies are conducted over extended periods of time, the adverse effects of internal validity threats are "absorbed" into the trend line.
   e. Probable Operating Threats to External Validity. The interactive effect of the treatment with selection, history, or maturation; multiple treatment interference, experimenter effects; and/or reactive or interactive effect of testing may operate, as subjects were not randomly selected or assigned.

f. Statistical Analysis is by trend line where differences in the slope of the line are tested. If there is a significant difference in the slope, then the independent variable can be cited as the probable source.
g. Comments:
   (1) The addition of a comparison group improves this design. To increase external validity, repeat the study within different contexts and conditions.
   (2) The chief weaknesses of the design is that takes a long time to conduct and as such, is subject to the influence of confounding variables.

3. The Single Subject Design
If a single individual is studied, a Single Subject Design (Kazdin, 2010) may be utilized. Single subject designs are widely used in social work, psychology, medicine, and other professions where the focus is on individual treatment, education, or progress. The Single Subject Design is a variation of the time series design. McMillan and Schumacher (2010, pp. 285-288) present two single subject designs: A-B or A-B-A.
   a. The A-B Design
      (1) In the A phase, there are multiple instances of data collection from the dependent variable to establish a baseline.
      (2) An intervention (independent variable) is applied in the B phase, with multiple instances of data collection from the dependent variable.
      (3) Figure 8.5 is an example, if an individual and not a group is involved; O_1 to O_4 is the baseline; O_6 to O_8 are the intervention data.

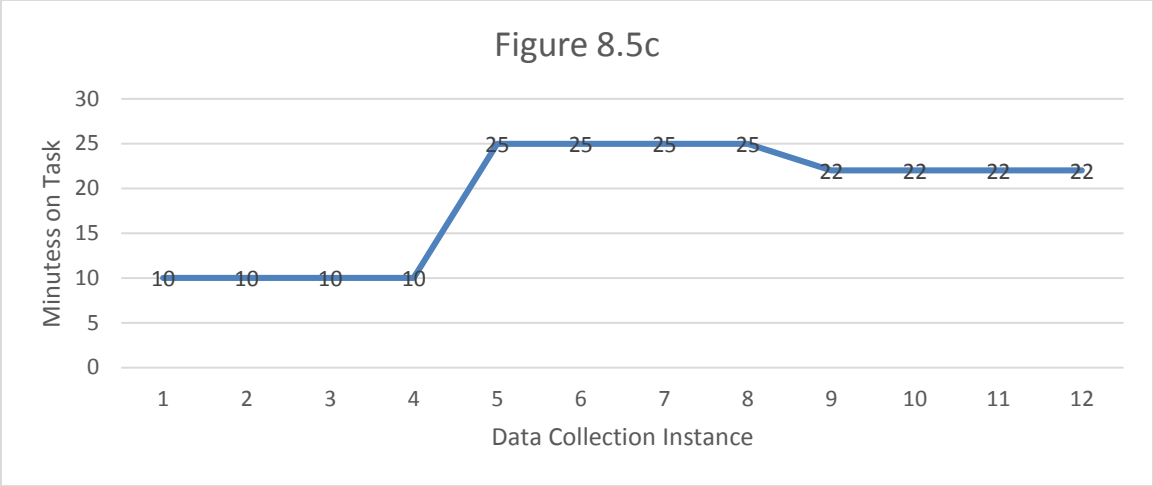
   b. The A-B-A Design
      (1) Phases A and B (The A-B design) are implemented.
      (2) Next, phase A is repeated to establish a "new" baseline. Extending Figure 8.5a, we'd have four more instances O_9 to O_{12} of data collection. See Figure 8.5b

      Figure 8.5b

      O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8 O_9 O_{10} O_{11} O_{12}

      Time

   c. Referring to Figure 8.5c, we see a fictitious example where a student exhibited an inability to stay focused on an assigned task. The professor collected a baseline measurement (O_1 to O_4, A); introduced an intervention, collecting data during the intervention phase (O_5 to O_8, B); and after withdrawing the intervention, collected additional data to establish a new baseline (O_9 to O_{12}, A). We see a big jump in attention span after the intervention was initiated and a slight decline after withdrawal. We assume the intervention caused the shift.



C. True Experimental Designs

The true-experimental designs are considered the "Gold Standard" in quantitative research. The random selection of subjects (i.e., probability sampling) into the study sample and the random assignment of subjects to either the treatment or control group controls internal design validity threats and permits generalization of findings back to the parent population as random selection ensures the sample is representative of its population. Of course, the representativeness of the sample should be shown by the researcher.

1. Figure 8.6 Pretest-Posttest Control Group Design
   a. Diagram

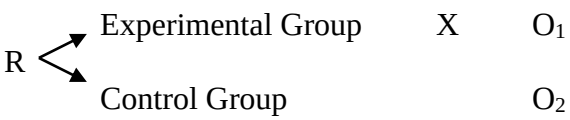
R →	Experimental Group	O_1	X	O_3
R →	Control Group	O_2		O_4

   b. <u>Probable Operating Threats to Internal Validity</u>. Resentful demoralization of the control group and diffusion of treatment may operate, if members of either group are "close" or particulars of the study are not kept confidential only as allowed by law or policy.
   c. <u>Probable Controlled Internal Validity Threats</u>. History, maturation, testing, instrumentation, regression, selection, mortality, and selection/maturation interaction are most likely controlled given the presence of random selection and assignment.
   d. <u>Possible Operating Threats to External Validity</u>. The interactive effects of the treatment with selection, instrumentation, history, or maturation; experimenter effects; reactive or interactive effect of testing; and/or reactive effects of the experimental arrangements may operate but any influence is greatly reduced given the presence of random selection and assignment.
   e. <u>Statistical Analysis</u>. Descriptive statistics should be routinely applied. Since two observations are taken, the two sample z-test or independent

samples t-test procedures are appropriate. If two or more groups are employed, the Analysis of Variance (ANOVA) or Analysis of Covariance (ANCOVA) procedures may be applied. The ANCOVA is particularly useful as it can statistically equate pretest group scores and can "control" for specific extraneous or moderating variables (Stevens, 1999, pp.307-308).

f. Comments
   (1) This is among the most reliable of the designs presented. It provides the "cleanest" evidence that the independent variable has exerted an effect on the dependent variable.
   (2) Randomization in subject selection and assignment and use of a control group ensures that the internal threats to validity are controlled
   (3) Generalizability may be reduced. However, if the sample is representative of the population and the treatment (independent variable) uncomplicated, then generalizability is most likely ensured. Replication of the study by others may confirm findings; thus demonstrating that the possible operating threats to external validity do not operate. To replicate, is one reason why full descriptions of experiments and other studies are so essential.

2. Figure 8.7 Posttest Only Control Group Design
   a. Diagram



   b. Probable Operating Threats to Internal Validity. Resentful demoralization of the control group and diffusion of treatment may operate if members of either group are "close" or particulars of the study are not kept confidential only as allowed by law or policy.
   c. Probable Controlled Internal Validity Threats. History, maturation, testing, instrumentation, regression, selection, mortality, and selection/maturation interaction are most likely controlled given the presence of random selection and assignment.
   d. Possible Operating Threats to External Validity. The interactive effects of the treatment with selection, history, or maturation; experimenter effects; reactive or interactive effect of testing; and/or reactive effects of the experimental arrangements may operate but any influence is greatly reduced given the presence of random selection and assignment.
   e. Statistical Analysis. Descriptive statistics should be routinely applied. Since two observations are taken, the two sample z-test or independent samples t-test procedures are appropriate. If two or more groups are employed, the Analysis of Variance (ANOVA) or Analysis of Covariance (ANCOVA) procedures may be applied. The ANCOVA is particularly

useful as it can statistically "control" for specific extraneous or mediating variables.

f. Comments
   (1) See above comments concerning Pretest-Posttest Control Group Design. This design is used when pre-testing is not possible.
   (2) Simplest statistical tests to apply are either a z-test or t-test or the ANOVA. The Analysis of Covariance procedure (ANCOVA) is preferred as it statistically equates the E- and C- groups.

D. Survey Research

Survey research is a very versatile data collection strategy; almost everybody has completed an online or paper-and-pencil questionnaire. Depending on the evaluation research question of the study, a purposeful or probability sampling strategy is used. Remember, survey research is a data collection strategy. A Questionnaire (or survey instrument) is a data collection tool. Be careful not to confuse the two. In this section, we first examine the survey research process or sequence and then the three specific survey data collection strategies. The interested reader is invited to review: Dillman, Smyth, and Christian, (2014); Fowler (2013); Nardi (2013); or Rea and Parker (2014) for more details.

1. Survey Research Process/Sequence
   a. Commencing the survey research project. The purpose, research questions or hypotheses, and scope (as influenced by purpose, budget and organizational capacity) are determined and should be reduced to writing.
   b. Survey design: The sampling frame is specified and constructed prior to the launching of a survey. Sampling frames may be based on either probability or non-probability sampling strategies. Also designed are the instrument, work-flow, data analysis plan, and reporting plan, nearly fully developed.
   c. Construct the data collection device. The actual measure is constructed using standard item writing guidelines. Take great care in item content, ordering, and presentation. The instrument should be as short as possible, well formatted, grammatically correct, etc. Ensure that directions and skip patterns are clear, complete, and accurate.
   d. Pre-test the data collection device. Select a group of pre-test subjects similar to intended participants. Administer the instrument in the same manner as will be done in the actual study. Instruct subjects to write comments about what the items meant to them, ease of following directions, recommendations for improving the instrument, etc.
   e. Revise the data collection device. Given the above, revise the instrument. If revision is substantial, repeat the pre-test. Continue this iterative process until satisfaction with the instrument is achieved. Once the instrument is finalized, the codebook to guide data entry and open-ended response coding is started.
   f. Data collection conducted. Data are collected, consistently, in the prescribed manner from the identified sample.

g Code preparation. The coding for closed-ended and simple open-ended items is fairly easy. However, for complex open-ended items, the codes "emerge" from the responses, as in content analysis. The researcher should review responses to these items, looking for common themes and code, based on those themes, so that similar responses are consistently coded in the same manner. See Appendix 4.7, Codebook.

h. Verification and editing. If the principal data collection strategies are personal or telephone interviewing, data verification is recommended. Every 10^{th}, 15^{th}, or 20^{th} subject is re-interviewed to ensure that the data reported are accurate. Editing is a process whereby responses are check for consistency and skip pattern conformance. If inconsistent responses are noted, the respondent can be contacted to clarify his or her response. Data verification and editing should be conducted during the research project. If telephone interviewing is done and data are directly entered into a database, the software will most likely conduct this phase.

i. Coding. At this stage open-ended items are coded. This should be done only by those who know the coding procedures as outlined in the codebook. Codes are recorded directly on the instrument in a response grid. If data are entered directly into a database, then this step is rarely necessary.

j. Entering data. A data entry operator enters data into a database with an automatic tabulation subroutine, spreadsheet (e.g., Excel) or a statistical program (e.g., SPSS or SAS). Data may be machine read by a scanner directly from a code response sheet. For other than small studies ($n \leq 200$ subjects) using a modest size instrument, scanning is highly recommended.

k. Tabulation and analysis. Once data are entered into a spreadsheet or statistical program, descriptive statistics and graphs of key variables are computed and studied. Next, cross-tabulations of key variables are computed. To analyze data, a person knowledgeable about the study, statistics, and the statistical software program is needed.

l. Recording and reporting. The report should be of sufficient length to summarize each phase outlined above. Data should be reported in a clear, concise, and accurate manner. The discussion section, if one is included, should draw on previous research to place the present findings within an interpretative context. Recommendations may be appropriate.

2. Primary Methodologies

a. The **directly administered** survey instrument is distributed by the researcher, postal service or through the Internet. Survey forms are self-administered and are of modest length. Self-administered instruments need to be reasonably short, thus limiting the amount and complexity of information which can be collected. Thus, efficient design is essential.

(1) Once the sampling frame is constructed, the instrument is fully developed and field-tested, the cover letter is written and data are collected.

(2) The cover letter should be personalized, well written, and signed including researcher title. McMillan and Schumacher (2010) have provided directions for writing the cover or reminder letters in mail surveys.
   (a) Identify the purpose of the study: Introduce the study, describing its purpose.
   (b) Importance of the study: Briefly describe the importance of the study and its benefits, including economic if any.
   (c) Importance of the respondent: Here the importance of responding is stressed. Request the potential respondent's cooperation.
   (d) Study endorsement: State any key endorsements of the study, names and titles of key officials and/or names of organizations. Keep the list brief and have proof of endorsement.
   (e) Describe the questionnaire: Briefly describe the measure in terms of time to complete, what the results will be used for, how the results will be reported, types of items, etc. Give a response deadline.
   (f) Assurance of confidentiality: State whether participation is voluntary and confidential or anonymous.
   (g) Closing: Thank the subject for responding. Provide a telephone number and an e-mail address for contact if needed.

(3) It is essential to maximize response rates. The key to accurately arguing that those responding are similar to those not responding is a high response rate. Those who are highly educated have an interest in the topic, or those who are angry tend to respond the most. For a survey sample to accurately reflect its population, a high response rate is a must. To increase response rates (usually to 70% or 80%) do the following:
   (a) In first mailing, include cover letter, survey form, and self-addressed, stamped envelope for return.
   (b) In the second mailing, send a gentle, professional letter about two weeks after first mailing.
   (c) In third mailing, send a cover letter, survey form, and self-addressed, stamped envelope for return about four weeks after the initial mailing.
   (d) In the fourth and final mailing, send a gentle, professional letter about six weeks after first mailing.

(4) Web-based surveys are growing in popularity and ease. However, comfort level with such studies is not high; this will increase. If a survey is "in-house" and all intended subjects have access to e-mail or the web, this might be a viable channel. There are very good quality web based software packages which will not only allow for collecting data, but will do some basic tabulation as well.

b. **Telephone Interviews**: a structured interview guide is completed by a trained interviewer with an eligible subject over the telephone.

(1) Telephone interviewers receive about 10-16 hours of training, depending on the length and complexity of the interview guide, where they become very familiar with the instrument and comfortable administering it via extensive role-playing. See the discussion on training within the discussion on personal interviewing.
(2) The sampling frame is usually constructed as the survey is conducted through random digit dialing, under rules of replacement and screening questions designed to identify eligible respondents. Random digit dialing is a process which generates telephone numbers based on telephone prefixes. Most of these generated telephone numbers are "dead air." The process is used as there are many unlisted telephone numbers and to secure a representative sample, everyone with a telephone has to have the opportunity to participate in the survey.
(3) Quality control is ensured either by monitoring calls or call backs to verify subject responses.
(4) When the desired sample size is achieved, the survey project is terminated.
(5) Telephone surveys are as expensive to conduct and manage as they are labor and technology intensive. However, through the use of this methodology, needed data can be gathered and transformed into information quickly.
(6) Conducting a telephone survey is becoming increasingly difficult with telemarketing, answering machines, computer modems, and voice mail. These devises tend to block connection or contact with possible subjects. However, if contact is made those owning these devises tend to participate, provided they are eligible.
(7) Computer Assisted Telephone Interviewing (CATI) has made paper and pencil interviewing less common. CATI allows the interviewer to collect data by keying responses onto the screen. When the interview is complete or at specified instances, data are dumped into a tabulation program which produces a report. CATI adds to the efficiency of telephone interviewing.
(8) Computer Administered Telephone Surveys (CATS) lack a human interviewer. Interviewing is conducted by computer using a synthesized voice. Refusal rates are higher.

c. **Personal or Face-to-face Interviewing**: a structured interview guide is completed by a trained interviewer in either an individual or group setting.
(1) Patton (2014) developed an interview taxonomy:
(a) Informal conversational interview: There are no predetermined topics or wording. Items emerge from the immediate context of the conversation. This approach is useful for exploring a subject's thoughts or feelings about a general topic or issue. Once several of these interviews have been conducted, the researcher is better prepared to construct more structured interview instruments.

(b) Interview guide approach: Topics and issues are specified in advance. The wording of items and item sequence is determined by the interviewer.
(c) Standardized open-ended interview: Topics, issues, items, and item sequence are predetermined and ordered in advance of the interview. All respondents are administered the same ordered instrument.
(d) Closed quantitative interviews: All items and response options are predetermined. Respondents select from these predetermined options.

(2) As the interviewing instrument becomes more structured, the role of the interviewer becomes less central and response variability is reduced. The lessening of the centrality of the interviewer will reduce interviewer bias. However, the subject is less likely to be spontaneous and/or detailed in his or her responses. Thus, the closed quantitative interview is not recommended.
(3) Interviewer training must be conducted and be very through.
(a) Training should include extensive study of the instrument, role-playing as interviewer and respondent, and practice of probing.
(b) Probing is a strategy which is designed to assist the respondent to fully answer an item. Probing should be gentle, consist of repeating the item, defining a term the respondent doesn't understand, etc. Often, repeating the subject's response will promote further comment and/or clarification. However, probing should not frame or influence a response for or by the subject.
(c) During training, the interviewers should practice establishing a comfortable interviewing environment; recording responses quickly and accurately; placing the subject at ease, and being at ease themselves; distracting or threatening body language should be identified and unlearned. Body language which promotes establishing a good relationship should be identified and practiced.
(d) The interviewer is the most influential person (aside from the subject) in a research project. Sufficient time should be spent in training. Closely supervise early interviews to ensure high quality data.
(4) Group-administered face-to-face interviewing is more cost and labor efficient than one-on-one interviewing. Once completed, the researcher has his or her completed instruments in hand and ready for tabulation and analysis.
(5) Interview Errors and Their Control
(a) Non-response Error. Intended subjects are not available or refuse to participate. This can introduce selection errors into any study. Non-response error can be lessened by employing one or more of these strategies:
(1) Call back. Attempt to contact the subject 3 or more times, by varying the time of day and day of week. This will usually

ensure that a substantial majority of intended subjects will have the opportunity to decide whether or not to participate in the study.

(2) Statistical weighting. Responses will be statistically treated so that responses from an under-sampled group are weighted to match population estimates.

(3) Substitution. Replace respondents who are not available. This is recommended only if replacements are known to match key characteristics of those failing to participate.

(b) Response Error. There are two sources for this type of error, the interviewer and respondent.

(1) Respondents fail, for whatever reason, to fully and accurately answer items. This failure is more likely when money and other sensitive content are involved.

(2) Interviewers fail to record answers fully and accurately, don't follow procedures, lie, falsify completed interviews, or unduly influence responses. To reduce the likelihood of the above, use highly trained closely supervised interviewers.

II. Qualitative Evaluation Research Designs

Qualitative research is a system of inquiry which seeks to build a holistic, largely narrative, description to inform the evaluator's understanding of a social or cultural phenomenon (e.g., context, event, etc.). Qualitative research takes place in natural settings employing a combination of participation, observations, interviews, and document or other artifact reviews. Qualitative research, as a research strategy, is based on underlying assumptions and perspectives. Wiersma (1995, pp. 211-212) summarized these as:

a. Phenomena are viewed in its entirety or holistically. It is not possible to reduce complex phenomena into a few interdependent or independent variables. Variables are not manipulated.
b. Investigators research in "nature." Researchers do not impose their assumptions, limitations, and delimitations or definitions, or research designs upon emerging data, as in quantitative research. The researcher's role is to record what he or she observes and/or collects from subjects' in their natural environment.
c. "Reality" exists as the subjects see it. The researcher is to record, fully, accurately and unbiasedly, that reality as seen through the eyes of subjects.
d. Conclusions emerge from the data; no *a priori* conclusions are made.

We will review the qualitative research process, and then six qualitative research strategies: ethnographic, phenomenological, and grounded theory, case study, focus group, and historical research. Qualitative research data collection strategies typically rely on a purposeful sampling strategy. There are several sources recommended for those interested in more detailed reading. The purpose of the present discussion is to

acquaint the reader with basic concepts and vocabulary. McMillan and Schumacher (2010) provide a thorough survey of the qualitative research process**.**

A. The "General" Qualitative Research Process

1. A working design is formulated. A preliminary research plan is drawn, but is intended to be flexible. Here the field sites are selected through purposeful sampling, given the study's purpose. The duration of fieldwork is estimated and other relevant operational issues (e.g. securing necessary resources, permissions, etc.) are addressed. There are two widely used approaches for formulating a working design: the funnel or the analytic induction approach.
   a. In the working design phase, the researcher has a very general research question or hypothesis which is used to select the initial research site, subjects, research operations, data to be collected, etc. Overtime, through consultation with colleagues, research reviews, and conversation with study sponsors or possible subjects (or those who are quite similar), the research question or hypothesis becomes increasingly focused. This process is repeated until the research question and operations (e.g., data collection, analysis, and interpretation) focus exclusively on the phenomena under study and produces "solid" conclusions. The process is referred to as the "funnel approach." Qualitative research is a gradual, iterative process.

   b. As an alternative to the funnel approach, a qualitative researcher may use the modified analytic induction approach. Wiersma (1995, p. 219) says the researcher starts with specific research question(s); identifies virtually all instances (or cases) of the phenomenon under investigation; and investigates each and every case, employing an iterative process where the research question or phenomenon explanation is revised until a comprehensive, descriptively rich narrative is produced complete with findings and conclusions as in the funnel approach.

   c. The funnel approach is more economical as usually only one or two research questions are selected. But, this precision comes at the risk of missing potentially relevant, important information as every potential instance of the phenomenon under study isn't investigated. The choice between approaches must be carefully considered.

2. Working hypotheses or research question(s) evolve. Using an inductive mode of inquiry, qualitative researchers, refrain from positing firm hypotheses or any hypotheses at all. General research questions are typically posed and as data collection and analyses proceed, more specific questions usually emerge. These more specific questions and/or hypotheses may be extended, deleted, or reframed as data collection and analysis continues. The objective of this process is a comprehensive, accurate description of the phenomena being investigated from the perspective of those who experienced it.

3. Organic, evolving data Collection is launched. The chief data collection devices are observation, interview, artifact (i.e., records, documents, etc.), oral histories, and specimen records (behavior recorded through observation). The qualitative researcher is advised to keep fairly detailed records of his or her thoughts, feelings, and behaviors while data are collected. It is important to determine whether or not the researcher is himself or herself a source of bias. These notes also contain changes in the work design and research questions or hypotheses.

4. Data collection, analysis and interpretation are iterative, continuing until senior researchers conclude sufficient data have been collected or it becomes apparent that continued data collection only reveals what is already known.
   a. McMillan and Schumacher (1993, p. 479) defined qualitative research as, "primarily an inductive process of organizing data into categories and identifying patterns (relationships) among categories." This definition implies that data and meaning emerge "organically" from the research context. Qualitative research data files are typically quite massive.
   b. Data must be organized and reduced for analysis and interpretation. Data are organized by coding. Relevant descriptions of behavior, statements, feelings, thoughts, etc. are identified and coded.
      (1) Wiersma (1995, p. 217) identifies three types of codes:
         (a) Setting or context codes. These codes describe the setting or descriptors of the phenomenon under study. Given that copious field notes are taken, codes for specific or regularly occurring characteristics contribute to efficient and effective field note production.
         (b) Perception codes. These codes are used to accurately record subjects' perception, understanding, feelings, and impact, etc. of relevant people, circumstances, events, or other things.
         (c) Process codes. It is a given in qualitative research that naturally occurring systems change. These codes are used to note event or process evolution and factors which cause or contribute to that evolution.
      (2) These codes need not be mutually exclusively and rarely are. The specific coding system employed by a researcher usually emerges as the iterative data analysis and interpretative process unfolds. The coding system employed by the qualitative researcher should be:
         (a) Comprehensive and tailored to his or her needs,
         (b) Accurate in recording what is being observed or reported, and
         (c) Useful in describing and enabling understanding of the phenomenon under study.
      (3) Coding can become quite complex, even for small scale evaluation or research projects, requiring the use of coding software. The novice qualitative evaluator/researcher should seek guidance on how to construct coding schemes and selecting software. These codes are the

basis for data analysis and interpretation. The triangulation of coding, initial impressions, tentative conclusions, etc. is an important part of this process (See the discussion of triangulation below.)

5. Write the qualitative study report. Once code analysis is complete, a report is written, having two objectives: (1) demonstrate the study's internal validity (rich description, trustworthiness, and interpretive validity) and (2) demonstrate the study's generalizability (comparability and translatability).
   a. Establish the Study's Internal Validity (i.e., creditability). A qualitative study's internal design (creditability) relies on a rich description, trustworthiness, and interpretive validity to validate study's design, findings, and conclusions. Interpretive validity relies on the study's rich description and trustworthiness.
      (1) Prepare a rich description. Since it's virtually impossible to control variables in "natural" settings, it is essential that a detailed, full (i.e., "rich") description of the research site, subjects, data collection tools, research procedures, researcher impressions and feelings, etc. be written to set a thorough, complete context for a reader.
      (2) A study's "trustworthiness" is increased when data points, analysis and conclusions are triangulated; subjects' perceptions are verified in a systematic manner; and the project's data chain of evidence is established (Gall, Borg, & Gall, 1996). Building trustworthiness uses these tools to verify data, starting with data collection and analysis:
         (a) Triangulation. The use of multiple data collection tools (qualitative and/or quantitative), data sources, analysts, etc. to establish the accuracy of the findings.
         (b) Member Checking. Research participants should review findings for accuracy and representativeness.
         (c) Chain of Evidence. The logical relationship between research questions, research procedures, raw data, and results should be such that a reasonably prudent person would arrive at the same or similar conclusions. Five strategies for establishing the data's chain of evidence are:
            [1] Outlier Analysis. Highly dissimilar cases should be examined and differences explained. This will contribute to strengthening the findings' integrity.
            [2] Pattern Matching. This is similar to the goal attainment methods for evaluating a project. Here, the perceived benefits of an intervention are matched against those found. If such are matched, then the argument for "trustworthiness" is enhanced.
            [3] Representativeness Check. This strategy is akin to monitoring used in survey research. An interview or artifact is reviewed to assess its representativeness as compared to other similar interviews or artifacts.

[4] Long-term Involvement. This is similar to trend analysis. If data are collected over the long-term, then situation specific influences are "canceled" out.

[5] Coding Check. Here, multiple researchers code the data and check for differences. Those differences are then resolved. A high level of agreement between coders is very desirable.

(3) Interpretive validity is the degree to which data interpretation and conclusions are considered accurate so as to be reflective of the subjects' or phenomenon's "reality." There are four dimensions to interpretive validity; the greater the degree of acceptance by other researchers, the more valid the original researcher's interpretation is perceived (Altheide & Johnson, 1994).

(a) Usefulness. Usefulness is determined by the extent the report informs and stimulates further research, i.e., other competent researchers view it as useful and credible.

(b) Contextual Completeness. This refers to a complete, thorough, and detailed (i.e., rich) description (usually in narrative form) of the report as seen by knowledgeable colleagues and end-users.

(c) Researcher Positioning. Qualitative researchers have been referred to as "data collection devices," given the role(s) of the researcher in qualitative strategies. Thus, the researcher must document his, her, or their direct and indirect effects on the research site(s), participants, etc. in the study's rich description.

(d) Reporting Style. This refers to the extent the research report authors' description is perceived as authentic and accurate, by colleagues (fellow researchers and practitioners), research subjects and other knowledgeable, interested readers.

b. Establish the Study's Qualitative Reliability and Generalizability

(1) Document that present findings are consistent with those of other investigators, conducting similar research. Draw on the procedures for establishing a study's trustworthiness. Combined, these strategies strengthen qualitative reliability arguments. The reliability indices discussed in Chapter 3, relate to quantitative data collection tools, not qualitative.

(2) The qualitative researcher does not share the same level of concern for generalizability as does the quantitative researcher. Qualitative external validity (i.e., generalizability) concerns itself with comparability (i.e., the ability of other researchers to extend knowledge based on the "richness and depth" of the description) and translatability (i.e., the extent to which other researchers understand the results, given the theory and procedures underlying the study.)

B. Specific Qualitative Evaluation Research Designs

1. Ethnographic Research Strategy

Ethnography is rooted in anthropology, ethnography involves the study of an intact group, logically defined, in its natural context for a sustained time interval. The researcher is typically an observer or a participant observer (Creswell, 2013).

a. Purpose. Goetz and LeCompte (1984, pp. 2-3) describe ethnography as, "[an] analytical description of social scenes and groups that recreate for the reader the shared beliefs, practices, artifacts, folk knowledge, and behaviors of those people." Great emphasis is given to the relationship between culture and behavior.

b. Process. Ethnographic research is very labor and time intensive, involving extensive fieldwork in a natural setting. Usually a general research question(s) is (are) identified. Once entry is gained and rapport (or trust) is established, the research questions are continually refined becoming more focused. It is not uncommon for the larger research question(s) to be segmented into more numerous, focused ones, evolving over time.

c. Data Collection. Ethnographic researchers use multiple data collection devices so that interpretations maybe grounded and triangulated. Leedy, (1997, p. 159) outlines three specific data collection devices:
   (1) Participant Observation. Here the researcher may participate in the phenomenon under study to varying degrees. Observation runs a continuum from detached observer to complete participant observer. The researcher must take great care to be as objective as possible. He or she is the single greatest source of bias within this strategy. The researcher will usually record his or her thoughts, feelings, etc. when developing field notes so as to guard against bias when interpreting the collected data.
   (2) Ethnographic Interviews. These types of interviews are usually purposeful, employing open-ended items so that the subject's reality and perceptions can be documented, understood, and interpreted.
   (3) Artifact Collection. This is a secondary data collection strategy which typically includes unofficial documents, official documents, objects, etc. which provide insight into the lives, experiences, and perceptions of subjects.

d. Data Analysis. Data analysis within ethnographic research occurs as data are collected. The researcher codes and classifies data (e.g., events, observations, quotes, etc.) into a meaningful taxonomy. New data are compared and contrasted to old so as to note patterns, etc. This iterative process continues until the researcher is able to make "assertions" which describe the participants' "reality" and perspectives.

e. Communicating Findings. Findings are reported in the form of research based assertions supported by analytical vignettes, interview quotes, and interpretative observations, all intended to present a holistic, rich description of the experiences and perceptions of participants.

f. LeCompte and Schensul (2010) provide guidance on designing and conduction ethnographic research as well as essential methods (Schensul & LeCompte, 2012a), specialized methods (2012b) and interpretation (2012c). Emerson, Fretz, and Shaw, (2011) educate on writing filed notes. O'Reilly (2012) provides a concise, general overview of ethnographic methods as does Murchison (2010). Kozinets (2009) provides guidance on conducting ethnographic research online.

2. Phenomenological Research

Using the phenomenological approach, the researcher develops an understanding of a subject's or subjects' "reality" however he, she, or they so perceive (Leedy, 1997, p. 161). In essence, this approach investigates an individual's or group's perception of reality as he or she constructs it. These realities may be expressed as an event, program, relationship, emotion, etc. Phenomenology is rooted in philosophy.

a. Purpose. Phenomenology seeks to understand a person's or persons' a perspectives as he, she, or they experience and understand an event, relationship, program, emotion, etc. (Leedy, 1997, p. 161). The researcher often has a significant personal interest in the phenomenon under study.

b. Process. Once a phenomenon is selected, the researcher engages in much the same process as used in ethnographic study.

c. Data Collection. Phenomenologists may study one subject, but typically 6-10, who is or are purposefully selected. These researchers rely on semi-structured in-depth interviews. The researcher and subject(s) must work rather closely together to collect data.

d. Data Analysis. The researcher(s) must search interview transcripts to locate "meaningful units" which are small bits of text which are independently able to convey meaning. Phenomenologists search for themes and patterns, not categories, by logically linking these "meaningful units."

e. Communicating Findings. Phenomenologists communicate findings through detailed narratives exploring themes and patterns which emerged from data analysis and reduction. These themes and patterns are then placed within the context of virtually all instances of the phenomenon under study.

f. For further reading, Moustakas (1994); Smith, Flowers, and Larkin (2009); Sokolowski (2009); van Manen (2014); or Vagle (2014) are recommended.

3. Grounded Theory Research

Grounded theory is a general sociological, qualitative research methodology used to build naturalistic (i.e., qualitative) theory (Strauss & Corbin, 1994, p. 275). Its primary objective is theory building.

a. Purpose. Using naturalistic iterative data collection and relationship analysis processes, researchers derive, from the data, a theory (Creswell, 2013). Conceptually dense theory is the expected outcome of the inquiry.
   (1) Two or more related categories (or concepts) give rise to a theory (McMillan & Schumacher, 2010), which is referred to as a proposition (Strauss & Corbin, 1990, p. 278).
   (2) Since several conceptual relationships (i.e., concepts) are required to define a theory, such theories are said to be conceptually dense, arising from the data.

b. Process. Using the iterative processes of data collection and analysis, relationships between concepts are continually identified and refined so as to enable theory development.

c. Data Collection. Grounded theorists employ the same data collection devices as do other qualitative researchers. The process is iterative with early data being compared and contrasted with "newer" data to refine, discard, generate, or extend questions, hypotheses, or conclusions.

d. Data Analysis. Using the iterative process between data collection and analysis within grounded theory, the researcher seeks to identify patterns of interaction between and among subjects (not necessarily individuals) by logically liking two or more data categories (i.e., similar topics sharing the same meaning). Strauss and Corbin (1990) explained the three major data coding strategies used in grounded theory research:
   (1) Open Coding. As the initial coding effort, data are deconstructed into the simplest form possible, examined for commonalities, and sorted into categories.
   (2) Axial Coding. As the intermediate step, data are reassembled based on logical connections between categories.
   (3) Selective Coding. At this third stage of coding, the "core" category is determined and the relationships between it and secondary categories are posited. Core and secondary category relationships are later validated. Categories needing further refinement or development are completed.

e. Communicating Findings. Strauss and Corbin (1990) write that to achieve "integration", the core category (or concept) is presented as a story line which becomes the lens through which all other categories are examined.

The relationships are compared to the data for validation, refinement, or discard.

f. For further reading, Bryant and Charmaz (2010), Charmaz (2014), Corbin and Strauss (2007), Goulding (2002), Okaty (2012), or Birks and Mills (2015) are recommended.

4. Case Study

a. Case studies are routinely employed in business, medicine, the helping professions, and law.
   (1) In a case study, a single person, program, event, process, institution, organization, social group is investigated within a specified time frame (Creswell, 2013) using field notes obtained through observation, participant-observation, and/or artifact analysis. Interpretive validity and trustworthiness must be established.
   (2) If one or more quantitative data collection strategies are included, the data collection tools must have the appropriate quantitative reliability and validity established (see Chapter 3), even if qualitative and quantitative data collection strategies and tools are used together.

b. The Case Study Process
   (1) Purpose. Case studies are constructed to richly describe, explain, or assess and evaluate a phenomenon [e.g., event, person, program, etc.] (Gall, Borg, & Gall, 1996, p. 549).

   (2) Process. The "case" is studied onsite within its natural context. The data gathering process is often interactive as the researcher or researchers associate with persons involved in the "case" under study.

   (3) Data Collection. Data is collected primarily by fieldwork, but secondary data collection (see Chapter 6) and artifact analysis is usually employed as well. It is important that the researcher(s) understand the phenomenon from the perspective of the participants.

   (4) Data Analysis. Case study data are analyzed in the qualitative tradition.
      (a) Gall et al. (1996) outlined three approaches to case data analysis:
         [1] Interpretational Analysis. When employing this strategy, the researcher is looking for patterns (threads, constructs, commonalities, etc.) within the data to explain the phenomenon.
         [2] Structural Analysis. Investigating patterns which may be found in conversations, text, activities, etc., with little or no explication as to pattern meaning.
         [3] Reflective Analysis. The description and evaluation of the studied phenomenon based on judgment and intuition by a highly qualified expert.

(b) If quantitative data are collected, use one or more of the appropriate methods from Chapters 9-13 to analyze these data separately. Interpret the qualitative and quantitative data together.

(5) Communicating Findings.
(1) The case narrative richly and fully reports the subject's perceptions about the phenomenon being investigated (Leedy, 1997, p. 158).
(2) According to Leedy, researchers using the reflective analysis strategy try to draw their readers into the participants' experiences by using emotive writings, poems, etc.
(3) Leedy goes on to point out that researchers using the other two analysis approaches tend to use an objective writing style and effectively use tables, figures, matrices, etc.
(6) For further reading, Farquhar (2012), Kazdin (2010), Remenyi, (2012), Woodside (2010) or Yin (2013) are recommended.

5. Focus Groups

Focus groups are panels, facilitated by a moderator, who meet for a specified time period to exchange perspectives, knowledge, and/or opinions on a particular topic. Groups are rarely more than a dozen members. Political science and marketing use this design.

a. Purpose. Focus groups are panels, facilitated by a moderator, who meet for a specified time period to exchange perspectives, knowledge, and/or opinions on a particular topic. Groups are rarely more than a dozen members. Focus groups are often sponsored by research, marketing, corporate, or political organizations. Focus group studies may use either a purposeful or probability sampling strategy.
(1) Focus groups
(a) Can quickly and cheaply identify core issues of a topic;
(b) Can observe reactions to a research question or product in an open forum;
(c) Can explore new or unexpected information or reactions in subsequent focus groups; and
(d) Enable subjects to respond in their own words and their emotional intensity can be measured.
(2) Focus Group Disadvantages
(a) Using a purposeful sampling seriously limits generalizability. If generalizability is necessary, use probability sampling.
(b) The convenience sampling strategy commonly used in focus groups may introduce bias into the research process. To counteract, ensure that group membership is representative of the population of interest.

b. Process. A topical interview guide is employed by the focus group leader, to manage the discussion and ensure that all subjects participate. Brinkman and Kvale (2014) offer guidance on crafting interview questions. Cooper

and Schindler (2013) recommend running separate groups when seeking data from subgroups of a population or larger group of interest. They argue that freer and more intense discussion will ensue. Locating members for focus groups is usually done informally, which could result in bias.

c. Data Collection. Focus groups may be conducted in person, over the telephone or Internet, or videoconferencing. It is common to record activity either on paper, tape, or video to enable accurate transcription and reporting. Using trained observers and standard reporting forms is also common. It is also routine practice to blend survey research methods into focus group data collection.

d. Data Analysis. Focus group data are analyzed in much the same manner as are case studies, qualitatively. If quantitative data are collected, use one or more of the appropriate methods from Chapters 9-13.

e. Communicating Findings. After data are analyzed and interpreted, reports are issued which is a blend of rich narrative and any quantitative data are presented in the usual manner.

f. For further reading, Barbour (2008); Edmunds (2000); Krueger and Casey (2014); Puchta and Potter (2004); or Stewart and Shamdasani (2014) are recommend.

6. Historical Research

Historical research relies on records, diaries, oral histories, photographs, and other artifacts to describe, analyze, and explain past events, philosophies, etc. The artifacts and records used are driven by the particular study and its research question(s). Historical research relies significantly on inductive, logical reasoning.

a. Purpose. Historical research relies on records, diaries, oral histories, photographs, and other artifacts to describe, analyze, and explain past events, philosophies, etc. The artifacts and records used are driven by the particular study and its research question(s). Historical research relies significantly on inductive, logical reasoning.

b. Process. While lacking highly defined methodological traditions, historiography is widely used. Essentially, it is a four step process, with considerable overlap:

(1) Identification of the Research Problem. Here is identified the reason for the research project and its corresponding, assumptions, questions or hypotheses. If hypotheses are posed, they are in the form of speculations as to reasons, processes, consequences, and/or characteristics of an event, issue, personage, circumstance, etc. under investigation. Any posited hypotheses should be based on accurate factual assumptions.

(2) Collection and Evaluation of Source Materials. Documents (e.g., books, newspapers, journals, letters, transcripts, etc.), artifacts (e.g., pictures, equipment, records, tapes, film, pottery, art works, etc.), information databases (hardcopy or electronic), and/or oral histories are examined. These documents, artifacts, databases, etc. are either primary or secondary sources:
   (a) Wiersma (1995, p. 234) defines a primary source as, "an original or first-hand account of the event or experience."
   (b) A secondary source is any account that is not primary. Primary sources are more desirable than secondary sources.

(3) Once evidence is collected, it must be examined. External and internal criticisms are essential components of the examination.
   (a) External Criticism. External criticism is used to determine the physical authenticity (i.e., validity) of the record, document, or artifact related to its investigation.
   (b) Internal Criticism. Historical researchers employ internal criticism to assess the degree, if any, of bias within the content of a record, document, or artifact. Internal criticism also seeks to ascertain the historical meaning and accuracy of an artifact's content.

c. Synthesis of Information. Once each document is authenticated, its contribution to the research underway must be determined. As the synthesis process evolves, it may be necessary for the researchers to reframe research questions or hypotheses given available data. It may be necessary to repeat the second or even first step in the process.

d. Analysis, Interpretation, and Formulating Conclusions. Historical research relies on inductive logical information analysis. At this step, conclusions are derived and hypotheses either supported or not. Alternative explanations should be "explained away." Also, the researcher should be as objective as possible. There should be substantial citations to support the author's (s') conclusions.

e. Data Collection. As noted above, data collection is a function of identifying documents, artifacts, etc.; examining their authenticity; chronologically ordering them; and then determining value or contribution to the research effort.

f. Data Analysis. Cross-referencing (triangulation) is essential to establish the veracity of a single document, artifact, etc. Each document, artifact, etc. needs to undergo chronological examination. Core ideas, concepts, and facts, need to be "pulled together" so as to make sense given the context of the period of time or event under study.

g. Communicating Findings. Reports of historical research are usually presented in narrative, absent headings; however, this is changing. The

purpose of the study is typically cited first and placed within a context. Next, presented are any research questions or hypotheses with supportive or contradictory evidence; the hypothesis is either retained or refuted. Once all the evidence is presented, conclusions are drawn.

h. For further reading, Brundage (2013), Bryant and Hall (2005), Danto (2008), or McDowell (2014) are recommended.

III. Specialized Designs

A. Correlational Research

Correlation research is a design where the researcher searches for relationships between and among variables.

1. Correlation coefficients are used. The discussion and associated references, presented in Chapter 11 of this primer, will provide the reader with a basic understanding of the zero-order correlation. Some are tempted to attribute causation, i.e., "X" causes "Y" based upon correlation studies. Imputing causation is technically incorrect.

2. Correlation research uses existing or newly constructed datasets, which are usually quite large. Advanced correlational research tools include path analysis, stepwise multiple regression, probit regression, logistic regression, canonical correlation, factor analysis, and cluster analysis. Since these analytical tools require significant advanced statistics reading and training; a consultant is recommend if the reader lacks the necessary knowledge and experience. Additional substantive discussions are found in Bobko (2001), Cohen, Cohen, West, and Aiken (2002) or Montgomery, Peck and Vining (2012).

B. Causal-Comparative or Ex-Post Facto Design

The Causal-comparative or Ex-post facto design enables a researcher to examine cause-and-effect relationship(s) where it would be illegal, impossible, or unethical to manipulate the independent variable(s).

1. Independent and dependent variables are identified within a sample drawn from a population of interest. The purpose of which is to search backward from consequent data (i.e., dependent variable) to antecedent causes (i.e., independent variable or variables.) Attribution is based on logical analysis and the "accumulated body of evidence."
2. A strategy for conducting ex-post facto research, is to identify a sample on a given dependent variable (e.g., heart disease) and then to sort (or divide) the sample into relevant sub-groups, by key independent variables (e.g., smoking, blood pressure status, exercise, etc.). It may be necessary to match participants or adjust statistically for differences between and among groups as necessary to ensure that the sub-groups are representative of the sample, except on the independent variable(s) of interest.
3. Causal-comparative research, like survey or correlation methods, will either employ hypotheses and/or research questions. Existing data-sets are

employed. Statistical procedures are routinely applied to data for analysis purposes.

4. Additional discussions of this design approach are presented in most introductory and intermediate research methods textbooks or may be found on the Web. While a powerful design, the statistical analysis of data can be quite complicated; the reader who is considering such a design is advised to consult a knowledgeable and experienced consultant.

Review Question Set I: Pre-, Quasi, & True- Experimental Designs

Directions. Read each item carefully; either fill-in-the-blank or circle letter associated with the term that best answers the item. These items relate to pre-experimental, quasi-experimental, or experimental evaluation designs.

1. Within a _____ study measurements of the dependent variable are conducted once.
   a. Longitudinal
   b. Ex post facto
   c. Cross-sectional
   d. Panel

2. Criteria for determining an adequate sample include:
   a. Representativeness
   b. Randomness
   c. Precision
   d. "a" and "c"

3. A research design in which cases are randomly assigned to groups which are tested or studied is:
   a. Quasi-experimental design
   b. Pre-experimental design
   c. Experimental design
   d. Survey design

4. The pre-experimental research design where only history, maturation, selection, and mortality are uncontrolled is:
   a. One-shot case study
   b. "Before and After"
   c. Static-Group comparison
   d. Two-shot case study

5. Which one of the following experimental designs is most efficient concerning resource consumption while still controlling threats to internal validity?
   a. Pretest-Posttest
   b. Solomon Four-group
   c. Post-test only
   d. Static-Group comparison

6. Which one of the following designs is most like the Pretest-Posttest experimental design?
   a. Time series
   b. Nonequivalent control group
   c. Counterbalanced
   d. Static-Group comparison

7. Which one threat to internal experimental validity is possibly not controlled by the Nonequivalent control group design?
   a. History
   b. Testing
   c. Regression
   d. Selection

8. _____ experimental designs are those where assertions about casual relationships can be made.
   a. True
   b. Quasi-
   c. Pre-
   d. Correlation

9. The basic assumption of the pretest-posttest control-group design is that the experimental and control groups are equivalent prior to the introduction of the
   a. Dependent variable
   b. Data analysis
   c. Moderating variable
   d. Independent variable

10. One critical element of any experimental design is:
    a. Subject selection procedures
    b. E- and c-group equivalence
    c. Implementation quality
    d. Data analysis strategies

11. An advantage of the post-test only design to external validity is its control of:
    a. Pretest sensitization
    b. Regression
    c. Posttest sensitization
    d. Instrumentation

12. The two critical differences between true and quasi-experimental designs are:
    a. No random assignment or selection
    b. No random assignment or control groups
    c. No random selection or control groups
    d. Generalizability and control groups

13. The one-shot case study:
    a. Has low internal validity
    b. Has no generalizability
    c. Controls virtually no internal validity threats
    d. All are correct

14. For the true experimental designs presented, the preferred statistical procedure is:
    a. t-test
    b. ANCOVA
    c. MANOVA
    d. ANOVA

15. Controls for non-response error includes all except:
    a. Substitution
    b. Increased sample size
    c. Statistical weighting
    d. Call back

Answers: 1. c, 2. d, 3. c, 4. a, 5. c, 6. b, 7. c, 8. a, 9. d, 10. c,11. a, 12. a, 13. d, & 14. b, 15. d.

Review Question Set II: Survey Research

Directions. Read each item carefully; either fill-in-the-blank or circle letter associated with the term that best answers the item. There are review items for survey research.

1. _____ is the most widely used research method in education, business, and other social sciences.
   a. Experimental research
   b. Quasi-experimental research
   c. Survey research
   d. Correlational research

2. Surveys are most commonly conducted for the purpose of
   a. Description
   b. Projection
   c. Exploration
   d. Detection

3. Which one of the following is not a survey research application?
   a. Face-to-face interviews
   b. Mail
   c. Self-administered
   d. Observation

4. _____ is a form of non-response associated with face-to-face interviewing.
   a. Accessibility
   b. Exclusion of groups
   c. Selection
   d. Maturation

5. To determine whether or not survey items can be consistently understood and answered, a researcher's best strategy is
   a. Conducts a pilot test
   b. Writes items using writing guidelines
   c. Works with a willing small group of respondents who provide assistance
   d. To self-administer the survey form and render a professional judgment

6. Which one of the following statements about causal-comparative research is incorrect?
   a. Independent and dependent variables are identified within a sample.
   b. The independent variable or variables are drawn from consequent data.
   c. Attribution is based on logical analysis and correlational procedures.
   d. Cause and effect attributions should be avoided.

7. Which statement about the differences between field and survey studies is incorrect according to Schwab?
   a. Field studies do not rely on probability sampling models.
   b. Field studies are simpler in design and scope than surveys.
   c. Field studies test conceptual models, typically with more independent variables.
   d. Surveys investigate few independent variables.

8. The statement, "Independent and dependent variables are measured two or more times on a group of cases [subjects],"describes which one survey research design?
   a. Descriptive
   b. Predictive
   c. Panel
   d. Explanatory

9. To maximize survey response rates a strategy has been recommended. Which step in this strategy is incorrect?
   a. In first mailing, include cover letter, survey form, and self-addressed, stamped envelope for return.
   b. In the second mailing, send a gentle, professional letter about two weeks after first mailing.
   c. In third mailing, send a cover letter, survey form, and self-addressed, stamped envelope for return about four weeks after the initial mailing.
   d. In the fourth and final mailing, send a gentle, professional letter about eight weeks after first mailing.

10. The type of interview where the topics and issues are specified in advance. The wording of items and item sequence is determined by the interviewer is the _____ interview:
    a. Informal conversational
    b. Interview guide approach
    c. Standardized open-ended
    d. Closed quantitative

11. Concerning telephone surveys, which statement is most incorrect?
    a. The sampling frame is constructed as the survey is conducted.
    b. Training involves extensive role playing and may last 16 hours.
    c. Random digit dialing is a method which includes unlisted telephone numbers.
    d. Telephone surveys tend to be expensive but easily managed.

12. Controls for non-response error includes all except:
    a. Substitution
    b. Increased sample size
    c. Statistical weighting
    d. Call back

Answers: 1. c, 2. a, 3. d, 4. a, 5. c, 6, b, 7. b, 8. c, 9. d, 10. b, 11. d, 12. b,

Review Question Set III: Qualitative Designs

Directions. Read each item carefully; either fill-in-the-blank or circle letter associated with the term that best answers the item. There are review items for qualitative research.

1. The research design which "involves the study of an intact group, logically defined, in its natural context for a sustained time interval. The researcher is typically an observer or a participant observer" is:
   a. Phenomenology
   b. Grounded theory
   c. Case study
   d. Ethnography

2. The qualitative research design, where the purpose of which is to understand participants' perspectives and view of social realities is:
   a. Phenomenology
   b. Grounded theory
   c. Historical
   d. Ethnography

3. The data collection strategies, "participant observation", "interviews", and "artifact analysis" are most frequently associated with _____ research:
   a. Case study
   b. Historical
   c. Ethnographic
   d. Grounded theory

4. In terms of methodology, which <u>one</u> of the following is most like ethnographic research?
   a. Case study
   b. Phenomenological
   c. Historical
   d. Focus group

5. In establishing the "trustworthiness" of a qualitative study, the strategy where "the perceived benefits of an intervention are matched against those found" is called:
   a. Chain of evidence
   b. Pattern matching
   c. Coding check
   d. Representativeness check

6. To establish the interpretive validity of a qualitative study, the strategy where "the extent of the report confirms and stimulates further research" is:
   a. Usefulness
   b. Research positioning
   c. Contextual completeness
   d. Reporting style

7. Which one of the following is not a common assumption shared by qualitative research traditions?
   a. Phenomena are viewed holistically.
   b. Investigators research in "nature."
   c. A priori conclusions are either confirmed or disconfirmed.
   d. A subject's perception is reality.

8. Regarding qualitative research data analysis, the coding strategy, which describes the "natural" setting is called:
   a. Context codes
   b. Process codes
   c. Perception codes
   d. Response codes

9. Which statement about qualitative research reliability is <u>incorrect</u>?
   a. Multiple raters or observers are recommended.
   b. If standardized rating scales or checklists are employed, compute equivalence reliability coefficients.
   c. Compare present project findings with those of prior researchers.
   d. Draw on procedures to establish "trustworthiness."

10. Which statement concerning qualitative research generalizability is incorrect?
    a. Qualitative generalizability concerns itself with comparability.
    b. Qualitative generalizability concerns itself with translatability.
    c. Is as critical to the qualitative as the quantitative researcher.
    d. Is related to the qualitative study's validity.

11. Qualitative research traditions have several common reference points. Which statement is not correct?
    a. Working design
    b. Variable manipulation
    c. Working hypothesis
    d. Massive data sets

12. The qualitative research tradition where the researcher must search transcripts to find "meaningful units" so that themes and patterns emerge is called ____ research
    a. Phenomenological
    b. Historical
    c. Grounded theory
    d. Case study

13. The purpose of triangulation is:
    a. To publish
    b. To provide validation information
    c. To provide more data
    d. To provide reliability information

14. The qualitative research design where the researcher "a single person, program, event, process, institution, organization, social group or phenomenon is investigated within a specified time frame, using a combination of appropriate data collection devices" most likely refers to:
    a. Phenomenology
    b. Ethnography
    c. Case study
    d. Historical

15. Regarding historical research, which <u>one</u> of the following statements is <u>incorrect</u>?
    a. Can be quantitative or qualitative depending on the purpose of the study
    b. Is used to explain the past and describe relationships between persons, places, and events.
    c. Is conducted in natural surroundings.
    d. Is very similar to quasi-experimental research.

16. The first step in historical research is
    a. Formulation of a research question
    b. Data collection
    c. Primary source identification
    d. Data abstraction

17. _____ refers to the accuracy of information obtained in historical research.
    a. External criticism
    b. Document analysis
    c. Internal criticism
    d. Data analysis

18. Primary sources of information include:
    a. Oral histories from eyewitnesses
    b. Reports from those who have read primary sources
    c. Books about the person, place or event under investigation
    d. "a" and "c"

19. Which one of the following statements is incorrect regarding case study research?
    a. Used to conduct a detailed study to understand problems and processes within the applicable context
    b. May involve more than one person or event
    c. Focus is on a single case which undergoes an in-depth analysis
    d. Each statement is correct.

20. Which one of the following is not a case study data analysis strategy?
    a. Interpretational analysis
    b. Reflective analysis
    c. Structural analysis
    d. Non-structural analysis

21. Characteristics of focus groups include all of the following except:
    a. Enable subjects to respond in their own words
    b. Enables subjects' emotional intensity to be measured
    c. Is easily generalizable
    d. Can quickly and inexpensively identify core topic issues

Answers: 1. d, 2. a, 3. c, 4. b, 5. b, 6. a, 7. c, 8. a, 9. b, 10. c, 11. b, 12. a, 13. b, 14. c, 15. d, 16. a, 17. c, 18. a, 19. d, 20. d, 21. C.

References

Altheide, D. L. & Johnson, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 485-499). Thousand Oaks, CA: Sage.

Barbour, R. (2008). *Doing focus groups*. Thousand Oaks, CA: Sage.

Best, J. W. & Kahn, J. V. (2006). *Research in education* (10th ed.). New York, NY: Pearson.

Birks, M. & Mills, J. (2015). *Grounded theory: A practical guide*. Thousand Oaks, CA: Sage.

Bobko, P. (2001). *Correlation and regression: Principals and applications for industrial/organizational psychology and management* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Bordens, K. & Abbott, B. B. (2010). *Research design and methods: A process approach* (8th ed.). Columbus, OH: The McGraw-Hill Companies.

Brinkman, S. & Kvale, S. (2014). *InterViews: Learning the craft of qualitative research interviewing* (3rd ed.). Thousand Oaks, CA: Sage.

Brundage, A. (2013). *Going to sources: A guide to historical research and writing* (5th ed.). Hoboken, NJ: Wiley-Blackwell

Bryant, A. & Charmaz, K. (2010). *The SAGE handbook of grounded theory*. Thousand Oaks, CA: Sage.

Bryant, J. M. & Hall, J. A. (2005). *Historical methods in the social sciences*. Thousand Oaks, CA: Sage.

Campbell, D. T. & Stanley, J. C. (1963*). Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Carey, M. A. & Asbury, J. (2012). *Focus group research*. Walnut Creek, CA: Left Coast Press.

Charmaz, K. (2014). *Conducting grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

Christensen, L. B. Johnson, R. B., Turner, L. (2010). *Research methods, design, and analysis* (11th ed.). Upper Saddle River, NJ: Prentice-Hall.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cooper, D. R. & Schindler, P. S. (2013). *Business research methods* (12th ed.). Boston, MA: McGraw-Hill Irwin.

Corbin, J. M. & Strauss, A. C. (2007). *Basics of qualitative research: Techniques and procedures for grounded theory*. Thousand Oaks, CA: Sage.

Couper, M P. (2008). *Designing effective web surveys*. New York, NY: Cambridge University Press.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Danto, E. A. (2008). *Historical research*. New York, NY: Oxford University Press.

Dillman, D A., Smyth, J. D. & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hobboken, NJ: John Wiley & Sons.

Edmunds, H. (2000). *Focus group research handbook*. Chicago, IL: American Marketing Association.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic field notes* (2nd ed.). Chicago, IL: The University of Chicago Press.

Farquhar, J. D. (2012). *Case study research for business*. Thousand Oaks, CA: Sage.

Fowler, F. J. (2013). *Survey research methods* (5th ed.). Thousand Oaks, CA: Sage.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996*). Educational research: An introduction*. White Plains, NY: Longman.

Goetz, J. P. & LeCompte, M. D. (1984). *Ethnology and qualitative design in educational research*. Lexington, MA: D. C. Heath.

Goulding, C. (2002). *Grounded theory*. Thousand Oaks, CA: Sage.

Kazdin, A. E. (2010). *Single-case research designs* (2nd ed.). New York, NY: Oxford University Press.

Kozinets, R. V. (2009). *Netnography: Doing ethnographic research online*. Thousand Oaks, CA: Sage.

Krueger, R. A. & Casey M. A. (2014). *Focus groups: A practical guide for applied research* (5th ed.). Thousand Oaks, CA: Sage.

LeCompte, M. D. & Schensul J. J. (2010). *Designing and conducting ethnographic research: An introduction* (2nd ed.). New York, NY: AltaMira Press.

Leedy, P. D.(1997). *Practical research: planning and design* (6th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Marsden, P. V. & Wright, J. D. (Eds.) (2010). *Handbook of survey research* (2nd ed.). Bingley, UK: Emerald Group Publishing.

Martella, R. C., Nelson, R. & Marchand-Martella, N. E., (1999). *Research methods: Learning to become a critical research consumer*. Boston, MA: Allyn & Bacon.

McDowell, W. H. (2014). *Historical research: A guide for writers of dissertations, theses, articles and books*. New York, NY: Routledge.

McMillan, J. H. & Schumacher, S. (1993). *Research in education: A conceptual understanding*. New York, NY: Harper Collins.

McMillan, J. H. & Schumacher, S. (2010). *Research in education: Evidence-based inquiry* (7th ed.). New York, NY: Pearson.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis*. Hoboken, NJ: John Wiley & Sons.

Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage.

Murchison, J. M. (2010). *Ethnography essentials: Designing, conducting, and presenting your research*. San Francisco: Jossey-Bass.

Nardi, P. M. (2013). *Doing survey research* (3rd ed.). Boulder, CO: Paradigm Publishers.

Okaty, J. S. (2012). *Grounded theory*. New York, NY: Oxford University Press.

O'Reilly, K. (2012). *Ethnographic methods* (2nd ed.). New York, NY: Routledge.

Patton, M. Q. (2014). *Qualitative research and evaluation methods: Integrating theory and practice* (4th ed.). Thousand Oaks, CA: Sage.

Punch, K. F. (2003). *Survey research: The basics*. Thousand Oaks, CA: Sage.

Puchta, C. & Potter, J. (2004). *Focus group practice*. Thousand Oaks, CA: Sage.

Rea, L. M. & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide*. San Francisco, CA: Jossey-Bass.

Remenyi, D. (2012). *Case study research: The quick guide series*. Reading, England, UK: Academic Publishing International.

Russell, B. & Purcell, J. (2009). *Online research essentials: Designing and implementing research*. San Francisco, CA: John Wiley & Sons.

Schensul J. J. & LeCompte, M. D. (2012a). *Essential ethnographic methods: A mixed methods approach* (2nd ed.). New York, NY: AltaMira Press.

Schensul J. J. & LeCompte M. D. (2012b). *Specialized ethnographic methods: A mixed methods approach* (2nd d.). New York, NY: AltaMira Press.

Schensul J. J. & LeCompte, M. D. (2012c). *Analysis and interpretation of ethnographic data: A Mixed methods approach* (2nd ed.). New York, NY: AltaMira Press.

Schonlau, M., Fricker, R. D., & Elliott, M. N. (2002). *Conducting research surveys via e-mail and the Web.* Westport, CT: Rand Media.

Schwab, D. P. (1999). *Research methods for organizational studies*. Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis*. Thousand Oaks, CA: Sage.

Sokolowski, R. (2009). *Introduction to phenomenology*. Thousand Oaks, CA: Sage.

Stewart, D. W. & Shamdasani, P. N. (2014). *Focus groups: Theory and practice* (3rd ed.). Thousand Oaks, CA: Sage.

Stevens, J. P. (1999) *Intermediate statistics: A modern approach* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Strauss, A. & Corbin, J. (1994). Grounded theory: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273-285). Thousand Oaks, CA: Sage.

Sue, V. M. & Ritter, L. A. (Eds.) (2007). *Conducting online surveys*. Thousand Oaks, CA: Sage.

van Manen, M. (2014). *Phenomenology of practice: Meaning-giving methods in phenomenological research and writing*. Walnut Creek, CA: Left Coast Press.

Vagle, M. D. (2014). *Crafting phenomenological research*. Walnut Creek, CA: Left Coast Press.

Wiersma , W. (1995). *Research methods in education: An introduction* (6th ed.). Boston, MA: Allyn and Bacon.

Woodside, A. G. (2010). *Case study research: Theory, methods, & practice*. Bingley, UK: Emerald Group Publishing.

Yin, R. K. (2013). *Case study research* (5th ed.). Thousand Oaks, CA: Sage.