Key Terms

- convenience sample
- generalizability
- purposive sample
- random assignment
- random sample
- reliability
- replicability
- sampling
- validity

Suggested Readings

- Babbie, E. (2007). *The practice of social research* (11th ed.). Belmont, CA: Thompson.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston: Pearson Education.
- Creswell, J. W. (2013). *Qualitative inquiry and research design* (3rd ed.). Thousand Oaks, CA: Sage.
- Delice, A. (2010). The sampling issues in quantitative research. *Educational Sciences: Theory & Practice, 10*(4), 2001–2018.
- Law, M., Stewart, D., Letts, L., Pollock, N., Bosch, J., & Westmorland, M. (1998). *Guidelines for critical review form: Qualitative studies.* Available online: http://www-fhs.mcmaster.ca/rehab/ebp/pdf/qualguidelines.pdf
- Leech, N. L., & Onwuegbuzie, A. J. (2007). Sampling designs in qualitative research: Making the sampling process more public. *The Qualitative Report, 12*(2), 238–254
- Leech, N. L., & Onwuegbuzie, A. J. (2010). Guidelines for conducting and reporting mixed research in the field of counseling and beyond. *Journal of Counseling and Development, 88*(1), 61–69.
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach.* Thousand Oaks, CA: Sage.
- McCotter, S. S. (2001, June). The journey of a beginning researcher. *The Qualitative Report, 6*(2). Available online: http://www.nova.edu/ssss/QR/QR6-2 /mccotter.html

Web Links

- Basic Business Research Methods http://managementhelp.org/businessresearch/index.htm
- Qual Page: Resources for Qualitative Research http://www.qualitativeresearch.uga.edu/QualPage/
- Social Research Methods http://www.socialresearchmethods.net/

8

How to Write Chapter Four, Results

(Continued)

(Continued)

However beautiful the strategy, you should occasionally look at the results.

—Sir Winston Churchill

If you have completed Chapter Three and are ready to write Chapter Four, this means that you have finished collecting all your research data—bravo! You are more than halfway finished with the thesis, so keep the momentum going (and the coffee brewing). This chapter will focus on how to write Chapter Four, Results, of the thesis. In Chapter Four, you will report the study's results (for quantitative studies) or findings (for qualitative studies); in doing so, you will apply what you have learned from your data collection and analysis. In essence, this is the meat of your thesis. After all the blood, sweat, tears, eye strain, and hair pulling—what did you find out?

Chapter Four, Results, is an essential component of the master's thesis because you will report the outcomes of the study. This means reporting the results of the data analysis for each variable, data collection method, or measurement instrument that was used in the study. Therefore, you should have already consulted with your chairperson for the data analysis methodology. On occasion, you may need to make adjustments to the analysis or do additional analysis due to participants dropping out, and so on, before reporting the results. Depending on the research design and questions, the presentation of the results can be in narrative, numerical, tabular/graphic format, or a mixture of them all. For example, if you collected quantitative (i.e., numerical) data, the results will be reported in statistical or tabular/graphic format. These results are reported in a straightforward manner, the writing style is technical, and they can be monotonous. If you collected qualitative (i.e., nonnumerical) data, the findings will be reported in narrative and sometimes tabular/graphic format. You will provide thick descriptions of the data to paint a narrative "picture" for the reader. Chapter Four needs to be written with sufficient detail for replicability purposes in case someone wants to verify the results. In addition, how you report and organize the results here will determine how you interpret and discuss them in Chapter Five.

Preparation and Organization

There are several tasks that need to be completed before you begin to write. First, Chapter Four will be organized parallel to Chapter Three. Thus, I highly recommend that you make any final revisions to Chapter Three before writing Chapter Four. Second, make sure that all the data have been organized and analyzed. This will make the writing process go much faster. Typically, I advise students to analyze data collected from every measurement instrument (e.g., survey, test, interview, observation) to ensure nothing is overlooked. Depending on the data collected, this can involve simple scoring procedures and applying statistical tests or coding data and finding emerging themes. Third, if you are still struggling with data analysis, seek help from your chairperson. She can show you the best way to analyze the data or refer you to someone else. Your program or institution may also offer help with data analysis using a software program such as the Statistical Package for the Social Sciences (SPSS) for quantitative data or software computer programs to help code and analyze qualitative data (e.g., NVivo, HyperRESEARCH, HyperTRANSCRIBE).

Chapter Four Sections

Once you have analyzed all the data, you can start to write Chapter Four. Chapter Four starts on a new page in the thesis. Remember that writing a master's thesis is like telling the "story" of your research study. In Chapter Four, you are telling the main events (in this case, main findings) of the research study. However, unlike the first three chapters, there are no predetermined sections except for a brief introduction. This is because the sections in Chapter Four are dependent on the research design, research questions, and the specific data that were collected. These vary from study to study. Although there are no predetermined sections, there are common organizational strategies that are used to report the results. Keep in mind that although the sections are written and discussed separately, they are intertwined, and collectively they form the results or findings of the study. Check with your chairperson for how he or she wants you to organize the sections in Chapter Four.

To guide you in writing Chapter Four, I will discuss how to report results for quantitative and qualitative data separately. The rationale for discussing them separately is that the data analysis and reporting procedures are very distinct. I will also provide examples from different types of studies adapted from former students' completed master's theses. You will notice that the examples are written in the past tense to indicate that the data have already been collected and analyzed.

Quantitative Data

If you collected quantitative data, I recommend that you organize Chapter Four by reporting the results from each measurement instrument into separate sections. For example, if you used three measurement instruments, such as a test, survey, and an observation checklist, you would report the results in three separate sections with a subheading for each. You also want to report the results in the same order that the measurement instruments appeared in Chapter Three. Remember to use the three parallel ladders strategy from Chapter 7 (see Figure 8.1 for a depiction of the three parallel ladders strategy for Chapters Three and Four). This will make it easier for you to write and less confusing for the reader.

When faced with a quantitative data set, researchers need a way to organize the data to interpret and explain the results to others. Otherwise, the process of reporting the raw data would be overwhelming.

There are two main ways to analyze and report quantitative data collected from a sample group—using descriptive or inferential statistics. The type of data collected and the research questions will determine how you should analyze and report the data from each measurement instrument. For example, if the data were collected from a survey, you would likely report the data using descriptive statistics. If the data were collected from an experimental study, you would likely report the data using descriptive and inferential statistics. As part of your master's program, I assume that you have already taken a research methods course with an introduction to statistics. This discussion will therefore be a review to focus on the statistical procedures that are commonly used in a master's thesis. I will discuss each type of statistical analysis and reporting separately and share some examples from students' completed theses.

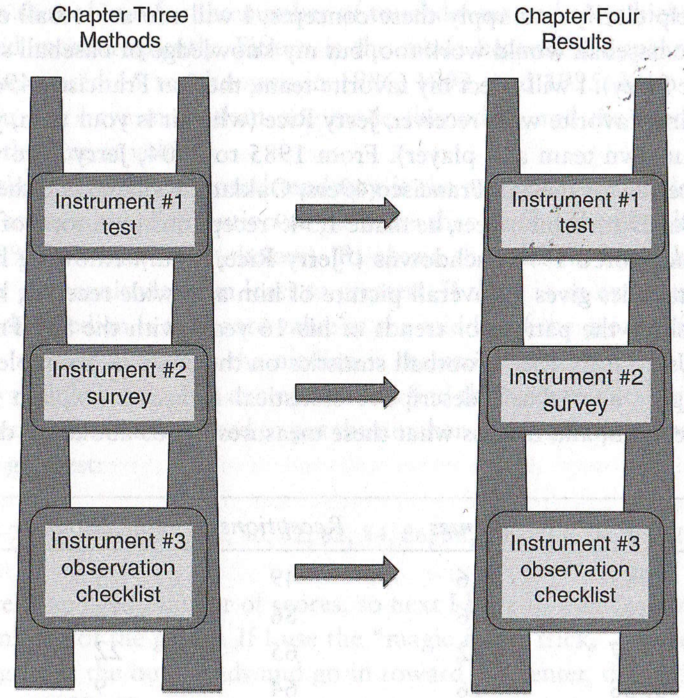


Figure 8.1. The three parallel ladders strategy for Chapters Three and Four.

Descriptive Statistics

Descriptive statistics refers to "a set of concepts and methods used in organizing, summarizing, tabulating, depicting, and describing collections of data" (Shavelson, 1996, p. 8). As the definition implies, researchers use this type of statistical analysis to *describe* the data set that was collected from the sample. Think of descriptive statistics as describing a picture of the quantitative results in a way that is comprehensible and meaningful for the reader.

Measures of central tendency. One major type of descriptive statistics is the measure of central tendency. The **measure of central tendency** is the "typical" or "average" score in a distribution. This is important because when you are looking at a large set of scores, there is too much information to digest. Knowing the typical or average score gives you a general sense of how the sample group fared. Usually, when someone says "average," I tend to think of the arithmetic mean. The mean is one type of measure of central tendency. The mode and the median are also measures of central tendency.

To help clarify and apply these concepts, I will use a football example (I'm sure baseball would work too, but my knowledge of baseball statistics is a little fuzzy). I will select my favorite team, the San Francisco 49ers, and my all-time favorite wide receiver, Jerry Rice (when it is your turn, you can pick your own team and player). From 1985 to 2004, Jerry Rice was the wide receiver for the San Francisco 49ers, Oakland Raiders, and the Seattle Seahawks. During his career, he made 1,549 receptions for a total of 22,895 yards and scored 197 touchdowns ("Jerry Rice," n.d.). Knowing his total career statistics gives an overall picture of him as a wide receiver, but they do not show the pattern or trends of his 16 years with the San Francisco 49ers. Using Jerry Rice's football statistics on the chart as a sample group, I will apply some basic descriptive statistical measures, explain how to calculate them, and discuss what these measures tell us about the data set.

Year	*Games*	*Receptions*	*Touchdowns*
1985	16	49	3
1986	16	86	15
1987	12	65	22
1988	16	64	9
1989	16	82	17
1990	16	100	13
1991	16	80	14
1992	16	84	10
1993	16	98	15
1994	16	112	13
1995	16	122	15
1996	16	108	8
1997	2	7	1
1998	16	82	9
1999	16	67	5
2000	16	75	7

The first measure of central tendency I will address is the mode. The **mode** is the most common or most frequently occurring score in the distribution. To obtain the mode, find the score that occurred most frequently. For example, if I look at the number of receptions that Jerry Rice made from 1985 to 2000, 82 would be the mode because it appeared two times (1989 and 1998), whereas all the other numbers appeared only once. How would you find the mode for the number of touchdowns scored? Simply go through the column for the number of touchdowns and see which number appeared most frequently. Fifteen is the mode because it appeared three times; he scored 15 touchdowns in 1986, 1993, and 1995. Keep in mind that you can have more than one mode (bimodal), and the mode is not always the largest value.

Another measure of central tendency is the median. The **median** is the middle score in the distribution or the score that divides the distribution in half (50% above and 50% below). To obtain the median, I put the scores in order of magnitude from least to greatest. If there is an even number of scores, the median is the score value in the middle of the group. If there is an odd number of scores, the median is the score value halfway between the two middle scores. For example, I will find the median number of Jerry Rice's receptions. First, I need to put the number of receptions in order from least to greatest:

7, 49, 64, 65, 67, 75, 80, 82, 82, 84, 86, 98, 100, 108, 112, 122

There is an even number of scores, so next I have to find the two scores in the middle of the group. If I use the "magic finger trick," where I point two fingers at the outer ends and go in toward the center, the two middle numbers are 82, 82.

7, 49, 64, 65, 67, 75, 80, **82**, **82**, 84, 86, 98, 100, 108, 112, 122

Normally, I would find the value that is halfway by adding the two middle numbers and dividing by two, but since they are the same, I know the median is 82. This means that 50% of the number of receptions that Jerry Rice made with the 49ers was above and below 82 (note that the mode and the median are the same for the number of receptions).

How would you find the median for the number of touchdowns Jerry Rice scored? First, put the number of touchdowns in order of magnitude from least to greatest. Then, since it is an even number of scores, use the magic finger trick to find the two scores in the middle of the group. Now, add those two scores together and divide them by two to get the halfway value. If you obtained a median value of 11.5 (10 + 13 divided by 2), you got it!

Finally, the most commonly used measure of central tendency is the mean. The **mean** is the arithmetic average and calculated by the sum of the scores divided by the number of scores in the distribution. For example, to find the mean number of Jerry Rice's receptions, first I add up all the receptions:

$$49 + 86 + 65 + 64 + 82 + 100 + 80 + 84 + 98 + 112 + 122 + 108 + 7 + 82 + 67 + 75 = 1{,}281$$

Then I divide the sum by the number of scores in the set (1,281 ÷ 16 = 80.06). Now I know that throughout his career with the San Francisco 49ers, Jerry Rice had a mean of 80 annual receptions.

How would you find the mean number of touchdowns scored per season? First, add up all the number of touchdowns. Then, divide the sum by the number of scores. If you obtained a mean value of 11 touchdowns (176 divided by 16), bravo!

Now that you know how to calculate measures of central tendency to confirm what a spectacular wide receiver Jerry Rice was, how would you apply them to data from your master's thesis? Basically, when you have a set of scores, you should report a measure of central tendency as part of your results to inform the reader about the average score. The scores can be for any variable (e.g., height, weight, achievement level, self-esteem, heart rates) and from a variety of sources such as tests, surveys, observation checklists, and so on. Typically, for the master's thesis, I recommend that students report the mean score because it is the most commonly used and takes into account every score in the data set. However, the mode and the median can also be appropriate (depending on the type of data that were collected).

Measures of variability. Knowing the measure of central tendency is important, but it does not give enough information about the data. For example, calculate the mean for each of the two groups of students' math test scores below.

Group A:	5	8	7	10	5
Group B:	8	1	5	14	7

The mean score for each group is 7. Based on this information, I could assume that the two groups of students did similarly well on the math test since they have the same mean score. Now, put the scores in order of magnitude:

Group A:	5	5	7	8	10
Group B:	1	5	7	8	14

If the maximum test score is 15, notice how the scores in Group A are closer together while the scores in Group B are more spread apart. There is not a huge difference in performance between the students in Group A, and they cluster closer to the mean; however, for Group B, there was one student who received a score of 1 and one student who received almost a

perfect test score. These scores are farther from the mean. With this information, I can see that the two groups are not very similar even though they have the same mean. Thus, knowing only the measure of central tendency (e.g., mean) is only part of the picture and can be misleading.

If you are describing a set of scores, you also need to report the measure of variability. A **measure of variability** indicates how close or spread apart (i.e., dispersed) the scores are in a distribution. In other words, how much do the scores differ from themselves and/or the mean of the distribution? If they differ quite a bit (scores are scattered), then there is a lot of variability. If they are pretty similar (scores are clustered), then there is less variability. There are many different kinds of measures of variability, but for the purpose of the thesis, I will discuss only the range and standard deviation since they are the most relevant.

The range is one measure of variability that you are probably already familiar with. The **range** is the difference between the largest and smallest scores in a distribution. You can calculate the range by subtracting the smallest score from the largest score. For Group A, the range is 10 – 5 = 5. What is the range for Group B? That's right. The range for Group B is 14 – 1 = 13. In comparing the two groups, Group B has a larger range, and the scores are more spread apart than Group A's scores. However, the range is of limited use because it only looks at two scores, the largest and smallest scores, and does not take into consideration the other scores in the distribution.

A more commonly used measure of variability is the elusive standard deviation. The **standard deviation** indicates how much the scores vary from the mean in a distribution. The formula for the standard deviation is the square root of the variance, which is the average squared deviation of each number from its mean. Huh? Don't worry—it is not critical for you to calculate the standard deviation by hand because most computer programs or calculators will do it for you (although I think you would enjoy it). However, it is important to understand what it means in interpreting the results. Basically, if the standard deviation is small, then the scores are closer to the mean. If the standard deviation is large, then the scores are more spread apart from the mean. For example, look at the two normal distributions on the graph in Figure 8.2. They both have a mean of 50, but Distribution A is tall and skinny with a standard deviation of 5 whereas Distribution B is short and wide with a standard deviation of 10. This means that the scores in the Distribution A are closer to the mean, and the scores in the Distribution B are more spread apart from the mean. If I had graphed the two earlier datasets of Groups A and B, it would be a similar picture with Group A as Distribution A and Group B as Distribution B.

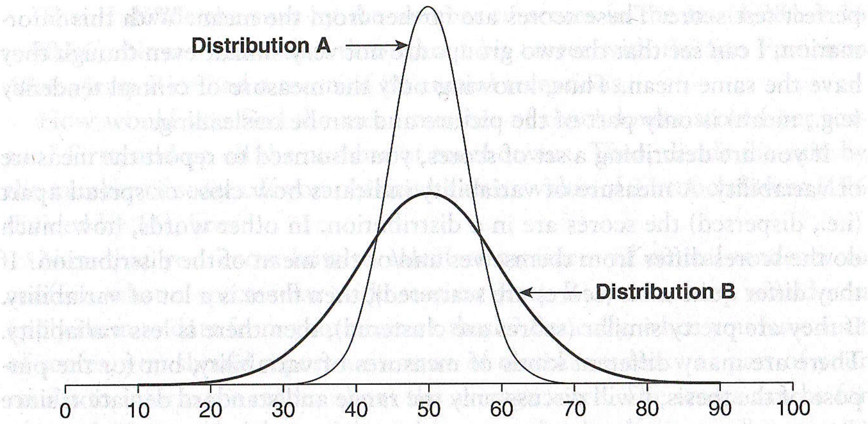


Figure 8.2. Normal distributions with different standard deviations.
SOURCE: Adapted from Lane, 2003.

Since the standard deviation is in relation to the mean, it is critical to report them together (you should also include the sample size). Within APA format, this can be done in several ways ("Reporting Statistics," n.d.). If you want to use abbreviations, they would be italicized, within parentheses, or at the end of the sentence. Here are the appropriate abbreviations to use: mean = M and standard deviation = SD. For example, "The 10 students in Group A had a higher mean score at the end of the intervention, $M = 18$, $SD = 2.3$." If there are two groups, then you can write, "The 10 students in Group A had a higher mean score ($M = 18$, $SD = 2.3$) than the 10 students in Group B ($M = 14$, $SD = 1.7$)." You can also write statistics spelled out as the subject of a sentence. For example, "The mean score on the math test for the 10 students in Group A was 18, and the standard deviation was 2.3." No matter which format you choose, remember to always report the sample size, mean score, and the standard deviation whenever possible.

Here is an example of results using descriptive statistics adapted from a former student's master's thesis (Henderson, 2007):

> The two measurement tools for social skill levels assessed students on their ability to perform 23 social skill tasks. These tasks ranged from making eye contact when speaking to someone to asking appropriately for help from an adult when needed. The first measurement tool was the teacher ranking survey in which the students' two teachers ranked the students individually on how well they were able to perform on each of the 23 social skill tasks. The only score generated from this survey was a total score. The range of scores for the teacher ranking survey was 36 (minimum 33, maximum 69). The mean total score of the teacher ranking survey was 52.29 points with a standard deviation of 10.13.

> The second measurement tool for social skill levels was the student self-rating questionnaire. Similar to the teacher ranking survey, the student self-rating questionnaire also assessed how well students could perform social skill tasks. However, the student self-rating questionnaire relied on the 14 students to rate their own ability to perform the tasks. The range of scores for the student self-rating questionnaire was 50 (minimum 64, maximum 114). The mean total score of the student self-rating questionnaire was 88.50 with a standard deviation of 13.24.

Additional ways to report data descriptively. In addition to measures of central tendency and variability, there are other ways to report quantitative data descriptively. This depends on the research questions and design as well as the intended message you want to convey to the reader. For example, you could report individual scores, percentages, frequency counts, and so on. I recommend that you include tables, charts, and figures as a graphical representation of the results to supplement the narrative explanation (see Chapter 10 for APA style).

If you have a study in which the sample group has only one participant, you would report the individual's scores. Here is an example of results for one participant adapted from a former student's master's thesis (Irey, 2008):

> Throughout the fluency intervention, Amber steadily increased the number of correct words per minute (CWPM) from baseline to phase III. During the baseline phase, Amber read 55, 60, 65, 63, and 58 CWPM, respectively ($M = 60$) (see Figure 1). During phase I, Repeated Reading, she read 64, 84, 73, 89, 89, and 84 CWPM, respectively ($M = 81$) (see Figure 2). During phase II, Error Correction, she read 85, 82, 74, 85, 78, 84, and 83 CWPM, respectively ($M = 82$) (see Figure 3). During phase III, Corrective Feedback, she read 76, 82, 83, 90, 87, 88, and 85 CWPM, respectively ($M = 84$) (see Figure 4). As indicated by the CWPM, Amber's mean reading rate greatly increased when she began Repeated Reading and increased slightly with the introduction of Error Correction and Corrective Feedback.

If you utilized a survey as a measurement instrument, you could report the frequency of responses in percentages across participants or for specific items. Here is an example of survey results adapted from a former student's master's thesis (Iniguez, 2007). In this study, the student administered the survey before and after the intervention. Therefore, she also reported the change in responses from her participants.

> The first survey item asked students about primary language instruction, "Being taught in Spanish at school makes me feel good about myself." On the preintervention survey, the mean was 4.09 ($SD = 1.37$). The frequency of

> the responses from the preintervention survey was: 9.1% chose "(4) A little like me," and 63.6% chose "(5) Totally like me." The postintervention survey results had a mean of 4.81 (SD = 0.40). The frequency of the responses from the postintervention survey included: 18.2% of the students chose "(4) A little like me," and 81.8% of the students chose "(5) Totally like me." The mean difference from pre- to postintervention survey was 0.72 and the response of "(5) Totally like me" increased by 18.2 percentage points.

If you observed participants' behaviors across multiple phases, you could report the individual or group data for each phase separately. Here is an example of frequency counts from a behavior intervention study adapted from a former student's master's thesis (Rau, 2006). In this study, the student had a baseline, treatment, and withdrawal phase. Therefore, he was able to compare the behaviors among the different phases and report changes in behaviors.

> Each type of off-task behavior was observed and recorded for the treatment phase. This information was used to determine whether there was an increase or decrease in behaviors from the baseline phase after the introduction of the Student Choice treatment.
>
> Incidents of *cross-talking* were observed and recorded for the treatment phase. Data indicated that there was a decrease in the range, total, and mean of cross-talking incidents from the baseline phase to the treatment phase. The range was eight, which was a decrease from 15. The total was 90 which was a decrease of 54 incidents. The mean number of incidents was 15, which was a decrease of nine incidents.
>
> *Total number of off-task behaviors* was observed and recorded for the *independent work time* of the treatment phase. Data indicated a decrease in the range, total, and mean number of total off-task behaviors observed during independent work time between the baseline phase and the treatment phase. The range of observed off-task behaviors during independent work time decreased from 21 in the baseline phase to seven in the treatment phase. The total number of incidents decreased from 204 to 117. The mean number of incidents decreased from 34 to 19.5.

Descriptive statistics are very useful to summarize, simplify, and describe the data in a study. However, they are also limiting because you cannot make any conclusions beyond the present data. For that I need to journey into inferential statistics. This would be a good time for that coffee break.

Inferential Statistics

Inferential statistics refers to "a set of methods to draw inferences about a large group of people from data available on only a representative subset of the group" (Shavelson, 1996, p. 8). In other words, researchers use sample group data to make assumptions or conclusions about the general population. This is very useful because most of the time researchers do not have access or the resources to collect data from the population. For example, consider how statistics are reported on presidential elections—how do they know that 46% will vote for Candidate A, 44% will vote for Candidate B, and 10% are undecided? Obviously, pollsters cannot ask every single person whom he will vote for in the next election. Instead, they ask a representative sample, apply statistical tests, and then make inferences about the rest of the country (remember, there is always a margin of error). Keep in mind that the sample must be representative (best done through random sampling); otherwise, the conclusions may be skewed toward one segment of the population or another. Basically, it is more realistic and efficient to collect data from a representative sample of the population to make inferences about the population rather than include the entire population in the study.

Tests of significance. Inferential statistics are also used in experimental studies. In these studies, tests of significance are conducted to determine if observed mean differences between groups or conditions represent a real difference or are due to chance. There are many different kinds of tests of significance, but for the purposes of the master's thesis, you would most likely not be required to go beyond applying a t test. A ***t* test** is a statistical test that is used to determine whether the observed differences between *two* mean scores represent a true difference or are due to chance. There are two different types of t tests: (a) independent-samples t test, and (b) nonindependent-samples t test (also referred to as dependent-samples t test or paired-samples t test). I will discuss each one separately.

Independent-samples *t* test. In a basic experimental study where one independent variable (cause) is manipulated to see its effect on one dependent variable (effect), the **independent-samples *t* test** is used to determine whether the difference in mean scores on the dependent variable between two independent groups is a real difference or one that is due to chance. In other words, is the mean score difference for the dependent variable due to the independent variable (treatment) or the result of some other chance factor such as sampling error? To use the independent-samples t test, the participants and their scores from the two groups must be completely independent and separate from each other.

For example, a researcher wants to determine if a math intervention (independent variable) will improve students' performance on the statewide math assessment (dependent variable). If the new math intervention results in significantly higher scores on the statewide math assessment, she will make a recommendation to the state education board to adopt the new math curriculum, so there is a lot at stake. The researcher randomly assigns 60 students into two groups: Group A gets the new math curriculum, and

Group B gets the traditional math curriculum. The students are exposed to the two treatments daily for eight weeks. At the end of the eight weeks, they all take the statewide assessment. The mean score on the statewide assessment for Group A was 90, and the mean score for Group B was 85. Since there is a five-point difference in favor of Group A, can the researcher make the recommendation to the state education board to adopt the new math curriculum? Not so fast. Unfortunately, researchers cannot simply eyeball the test scores and say, "Yes, five points seems like a big enough difference." You see, in statistics (and life in general), there is always room for error. Therefore, the researcher does not know whether the five-point difference represents a real difference (due to the new math treatment) or one that is due to chance. This is where the independent-samples *t* test comes in handy.

With this test, the researcher can determine the *probability* of whether the observed five-point mean difference between the two groups is statistically significant (i.e., represents a real difference). First, she needs to set up a null hypothesis (sorry, I was hoping to avoid this). The **null hypothesis,** H_0, represents the "chance" theory, meaning any observed differences are due to chance, and the treatment has no significant effect on the dependent variable. For example, the null hypothesis for the study would be as follows:

> H_0: There is no significant difference on the statewide math assessment scores between students who received the new math curriculum and the students who received the traditional math curriculum.

She can either reject or retain the H_0; typically, researchers want to reject the H_0 to "support" their new intervention. However, retaining the H_0 may be as valuable to the research literature as rejecting it (you may have discovered what treatment is not effective!). Remember that as the researcher, you are committed to reporting the findings objectively and accurately whether or not the data support your hypothesis.

Next, to determine whether or not to reject or retain the H_0, the researcher needs to set the probability or significance level (referred to as alpha, or α). The setting of the probability level is a bit like gambling, where the researcher gets to decide how much risk of making an error she is willing to accept. In social science studies, most researchers set the significance level at .05 ($\alpha = .05$), which means they are willing to take a 5% chance of making a Type I error. A **Type I error** is when you reject the H_0 when it is true. In other words, there is a 5% probability that the researcher concludes that the mean difference was due to the treatment when it was really due to chance. The good news is that she has a 95% of being correct (rejecting the H_0 when it is false)! After setting the significance level, the researcher conducts the independent-samples *t* test and compares the probability value

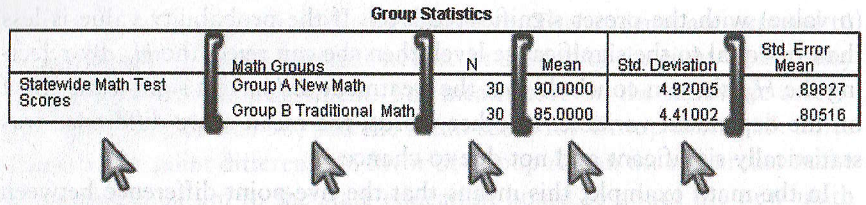
(*p* value) with the preset significance level. If the probability value is less than or equal to the significance level, then she can reject the H_0. By rejecting the H_0, she can conclude that the treatment *did* have a significant effect on the dependent variable. In other words, the mean score difference was statistically significant and not due to chance.

In the math example, this means that the five-point difference between the two groups was due to the treatment of the new math curriculum. The researcher could then make the recommendation to the state education board to adopt the new math curriculum. If the probability value is greater than the significance level, then she retains the H_0. By retaining the H_0, she concludes that the new math curriculum treatment did not have a significant effect on the statewide assessment scores, and the five-point mean difference was due to chance. The researcher should not recommend the new math curriculum to the state education board.

While it is not important for you to be able to conduct the independent-samples *t* test by hand (and this is not a statistics book), it is critical for you to understand its importance in determining cause-effect relationships in research studies and how to report these results. Since it is highly likely that you or your chairperson will use a computer software program such as SPSS to conduct the independent-samples *t* test, I will briefly discuss how to interpret the output and report the results in APA format. If you would like more information about the independent-samples *t* test or other significance tests, I highly recommend taking an introductory statistics course or perusing a statistics textbook.

Independent-samples *t* test SPSS output. When you conduct an independent-samples *t* test on SPSS, the first output screen you will see is the *Group Statistics* (see Figure 8.3 for group statistics output for independent-samples *t* test). These are the descriptive statistics for the total data set. There are several items to note in Figure 8.3. On the far left-hand side above the first arrow, you see *Statewide Math Test Scores.* This indicates the dependent variable. Moving to the next arrow between the first set of brackets is the two levels (groups) of the independent variable—*Group A New Math* and *Group B Traditional Math.* Above the next arrow is *N*, or the sample size, which is 30 in each group. The next two arrows between the second set of brackets indicate the mean scores and standard deviations for each group. Group A's mean is 90 (*SD* = 4.92), and Group B's mean is 85 (*SD* = 4.41).

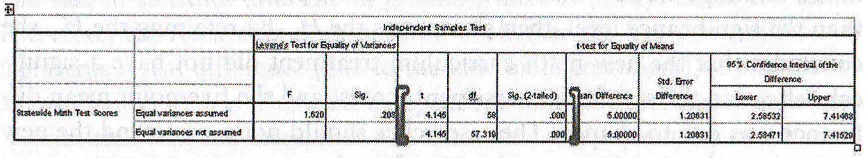
The next output screen you see is the *Independent Samples Test* (see Figure 8.4 for independent-samples *t* test SPSS output). These are the inferential statistics for the test. There are several statistics here, but I will only comment on the ones in the brackets since these are the items you will need to interpret and report in the results section of the thesis. Figure 8.5 is an enlarged section of the independent-samples *t* test SPSS



Group Statistics

	Math Groups	N	Mean	Std. Deviation	Std. Error Mean
Statewide Math Test Scores	Group A New Math	30	90.0000	4.92005	.89827
	Group B Traditional Math	30	85.0000	4.41002	.80516

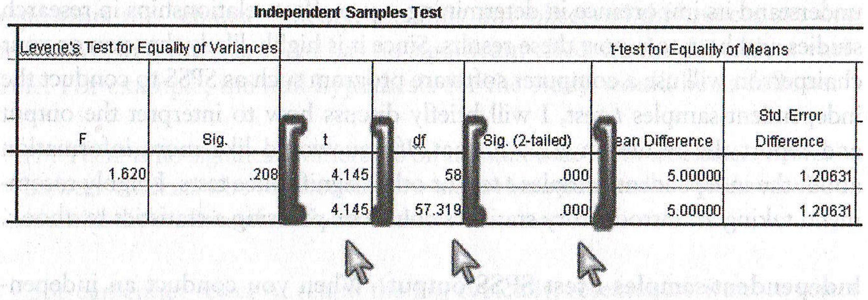
Figure 8.3. Group statistics from SPSS output for independent-samples *t* test.



Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	ean Difference	Std. Error Difference	Lower	Upper
Statewide Math Test Scores	Equal variances assumed	1.620	.208	4.145	58	.000	5.00000	1.20631	2.58532	7.41468
	Equal variances not assumed			4.145	57.319	.000	5.00000	1.20631	2.58471	7.41529

Figure 8.4. Independent-samples *t* test SPSS output.



Independent Samples Test

Levene's Test for Equality of Variances		t-test for Equality of Means				
F	Sig.	t	df	Sig. (2-tailed)	ean Difference	Std. Error Difference
1.620	.208	4.145	58	.000	5.00000	1.20631
		4.145	57.319	.000	5.00000	1.20631

Figure 8.5. Enlarged image of the independent-samples *t* test SPSS output.

output that I will focus on. On the left-hand side above the first arrow, you see the *t*. This *number* indicates the value of the *t* test. In Figure 8.5, $t = 4.145$. Alone, this does not tell us whether or not to reject or retain the H_0. Moving to the next arrow, you see *df*. This stands for degrees of freedom. For the independent-samples *t* test, the **degrees of freedom** is calculated by adding the sample sizes of the two groups together minus two ($n1 + n2 - 2$). In Figure 8.5, $df = 30 + 30 - 2 = 58$. Finally, in between the next set of brackets you see *Sig. (2-tailed)*. This is very critical and indicates the probability (*p* value). Remember the significance level of .05 that was set before the test was conducted? I need to compare the set significance level ($\alpha = .05$) with the probability value to determine if there is a statistically significant difference between the two groups. If the probability value is equal to or less than .05 ($p \leq .05$), then I can reject the H_0 and conclude that there is a significant difference between the two groups. If the probability value is greater than .05 ($p > .05$), then I retain the H_0 and conclude that there is not a significant difference between the two groups. In Figure 8.5, the probability is .000, which is definitely less than .05. YIPPEE! For researchers, getting a probability value less than .05 feels like winning the lottery (well, almost). In our study example, this means that the researcher can reject the H_0 and conclude that the new math curriculum treatment had an effect on the statewide math assessment scores. There was a statistically significant (real) difference in the mean scores between the new and traditional math groups that was not due to chance. She can make her recommendation to the education board to adopt the new math curriculum.

There are several variations of how you can report the results of an independent-samples *t* test in APA format. Be sure to include the two mean scores with standard deviations, *t* value with degrees of freedom, and the probability value. You should also report an effect size (usually in the form of Cohen's *d*) which represents the magnitude of the mean difference. If you have a small effect size, the two group's mean scores are not that different. If you have a large effect size, then the mean scores are very different from each other. Here is one example:

> To test the efficacy of the new math curriculum, an independent samples *t* test was conducted. This test was found to be statistically significant, $t(58) = 4.15$, $p < .001$, $d = .4$. The results indicated that on average, the students in the new math curriculum group ($M = 90$, $SD = 4.92$) performed better than the traditional math curriculum group ($M = 85$, $SD = 4.41$) on the statewide math assessment.

The independent-samples *t* test is the most common and simplest test to use when comparing mean differences between two independent groups. However, for a master's-level research study, it may be difficult to have access to a large sample with two separate and independent groups. More commonly, you may have access to only one group. For this type of research design, you need to utilize the nonindependent-samples *t* test. I will refer to this as the paired-samples *t* test since this is what is used in the SPSS computer software program.

Paired-samples *t* test. In a basic experimental study where one independent variable (cause) is manipulated to see its effect on one dependent variable (effect), the **paired-samples *t* test** is used to determine whether the difference in mean scores on the dependent variable between *two sets of related scores* is a real difference or one that is due to chance. This analysis is similar to the independent-samples *t* test except that with the paired-samples *t* test, there is no control group, and the scores are systematically related to each other. There are different ways for scores to be related, but typically the two sets of scores are from one group of participants. For example, a researcher wants to determine if a new reading intervention that uses high school English learners' primary language will enhance their vocabulary performance. He randomly selects one class of 30 high school students who are English learners. The students are given a pretest to measure their vocabulary level before the intervention begins. Then they receive the new reading intervention daily for 10 weeks. At the end of the 10 weeks, he administers the same test to measure their vocabulary level as a posttest. The pretest mean score on the vocabulary measure was 86 ($SD = 3.86$), and the posttest mean score on the vocabulary measure was 88 ($SD = 3.04$). Since the two sets of scores (pretest and posttest) are from the same set of students, they are in fact related.

In this scenario, the researcher needs to determine whether the two-point mean score difference between the pretest and posttest indicates a statistically significant difference (related to the vocabulary intervention) or one that is due to chance. The null hypothesis for the study would be as follows:

H_0: There is no significant difference between the high school English learners' pretest and posttest vocabulary mean scores.

To determine whether to reject or retain the H_0, he must set the significance level ($\alpha = .05$) and conduct a paired-samples t test. Since it is highly likely that you or your chairperson will use SPSS to conduct the test, I will briefly discuss how to interpret the output and report the results in APA format.

Paired-samples *t* test SPSS output. When you conduct a paired-samples t test on SPSS, the first output screen is the *Paired Samples Statistics* (see Figure 8.6 for paired-samples statistics output for paired-samples t test). These are the descriptive statistics for the total data set. There are several items to note in Figure 8.6. On the far left-hand side above the first arrow, you see *Pair 1*. This indicates the pair of scores that were compared—the vocabulary pretest and posttest scores. Moving to the next arrow between the first set of brackets are the mean scores for the pretest and posttest. The pretest mean score is 86.07, and the posttest mean score is 88.23. Above the next arrow is N, which is the sample size in each set of scores (note that the 30 for each group represents 30 scores for the pretest and 30 scores for the posttest, not 60 participants). The next arrow between the second set of brackets indicates the standard deviations for each set of scores, which are 3.86 and 3.04, respectively. The *Paired Samples Statistics* indicates that there is a two-point difference between the pretest

and posttest mean scores; however, we do not know whether this is a statistically significant difference or one due to chance.

The next output screen is the *Paired Samples Correlations*. These statistics tell you the correlation or the relationship between the two sets of scores (see Figure 8.7 for paired-samples correlations output). Above the first arrow in Figure 8.7, the *Correlation* column tells you the strength and direction of the relationship. The symbol to represent the correlation value is the letter r. In our example, the correlation between the two sets of scores is $r = .623$. This means that the students who had high scores on the vocabulary pretest also tended to have high scores on the vocabulary posttest and vice versa. Above the second arrow, the *Sig.* column tells you whether the relationship is statistically significant (not due to chance). In our example, the correlation is significant since p is less than .05 ($p < .001$).

The next output screen you will see is the *Paired Samples Test* (see Figure 8.8 for paired-samples test output). These are the inferential statistics for the test and similar to the ones for the independent-samples t test. There are several statistics here, but I will only comment on the ones on the right side in between the brackets since these are the items you will need to interpret and report in the results section. Figure 8.9 is an enlarged section of

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pretest Vocabulary Scores	86.0667	30	3.85901	.70456
	Posttest Vocabulary Scores	88.2333	30	3.03637	.55436

Figure 8.6. Paired-samples statistics SPSS output for paired-samples t test.

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Pretest Vocabulary Scores & Posttest Vocabulary Scores	30	.623	.000

Figure 8.7. Paired-samples t test correlations SPSS output.

Paired Samples Test									
		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Pretest Vocabulary Scores - Posttest Vocabulary Scores	-2.16667	3.08993	.56341	-3.31897	-1.01436	-3.846	29	.001

Figure 8.8. Paired-samples *t* test SPSS output.

Paired Samples Test

Paired Differences					
	95% Confidence Interval of the Difference				
Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
.56341	-3.31897	-1.01436	-3.846	29	.001

Figure 8.9. Enlarged paired-samples *t* test SPSS output.

the paired-samples *t* test output that I will focus on. Above the first arrow between the first set of brackets is the *t*. This indicates the value of the *t* score. In our example, $t = -3.846$. Notice that this time the *t* is a negative value. This indicates that the mean of the first set of scores (pretest) was less than the mean of the second set of scores (posttest). Moving to the second arrow, you see *df*, degrees of freedom. For the **paired-samples *t* test**, the degrees of freedom are calculated by the sample size minus one ($N - 1$). In Figure 8.9, $df = 30 - 1 = 29$. Finally, in between the next set of brackets above the third arrow is *Sig. (2-tailed)*. This indicates the probability of whether or not there is a statistically significant difference between the two sets of scores. Remember that I need to compare this to the set significance level of $\alpha = .05$. If the probability is equal to or less than $\alpha = .05$ ($p \leq .05$), I can reject the H_0 and conclude that there is a significant difference between the pretest and posttest mean scores. If the probability is greater than .05 ($p > .05$), then I retain the H_0 and conclude that there is not a significant difference. In Figure 8.9, the probability is .001, which is less than .05. YIPPEE! This means that I can reject the H_0 and conclude that the vocabulary intervention did result in a statistically significant (real) difference between the pretest and posttest vocabulary mean scores, and it was not due to chance.

To report this in APA format, be sure to include the two mean scores with standard deviations, *t* value with degrees of freedom, the probability value, and an effect size. Here is one example:

To test the efficacy of the vocabulary intervention for high school English learners, a paired samples *t* test was conducted. This test was found to be statistically significant, $t(29) = -3.846$, p = .001, $d = .5$. The results indicated that on average, the high school English learners scored significantly greater on the posttest vocabulary test ($M = 88.23$, $SD = 3.03$) than on the pretest vocabulary test ($M = 86.07$, $SD = 3.86$).

Here is an example of the results of a paired-samples *t* test adapted from a former student's master's thesis (Williams, 2006):

To analyze the results of the Arc's Self-Determination Scale (ASDS), first descriptive statistics were calculated for the pretest and posttest scores for each subgroup of the ASDS domains. The pretest mean scores and standard deviations for each subgroup of the ASDS domains were as follows: Autonomy ($M = .51$, $SD = .20$), Psychological Empowerment ($M = .73$, $SD = .15$), Self-Realization ($M = .75$, $SD = .16$), and Self-Determination Total ($M = .57$, $SD = .15$). The posttest means and standard deviations for each subgroup of the ASDS domains were as follows: Autonomy ($M = .63$, $SD = .18$), Psychological Empowerment ($M = .88$, $SD = .10$), Self-Realization ($M = .81$, $SD = .18$), and Self-Determination Total ($M = .68$, $SD = .14$). There were mean gains of .12 in Autonomy, .15 in Psychological Empowerment, .06 in Self-Realization, and .11 in the Self-Determination Total from pretest to posttest results.

Next, a paired-samples *t* test was conducted to determine if there was a significant difference between the pretest and posttest mean scores for each domain. There was a significant difference between the Psychological Empowerment pretest mean of .73 ($SD = .15$) and posttest mean of .88 ($SD = .10$), $t(10) = -3.16$, $p = .01$, $d = .3$, in favor of the posttest. There were no significant differences between the pretest and posttest mean scores for the other domains or total score.

In summary, the *t* tests for independent-samples and paired-samples are essential statistical tests to conduct when trying to determine whether the difference between two mean scores is statistically significant. They are easy to conduct using SPSS (or even by hand), and the output is straightforward for interpreting and reporting the results in APA format. In addition to the narrative explanation, it is also helpful to include tables or figures as part of the results. Now that I have discussed how to report the results of quantitative data, for you adventurous types, I will trek into our discussion on how to report the findings from qualitative data.

Qualitative Data

There is often a misconception among graduate students that interpreting and reporting data from qualitative studies is "easier" or "faster" than quantitative studies since there are no scary statistics involved. However, this is typically not the case. At the end of a qualitative study, a researcher may be faced with piles of data in the form of field notes from observations, transcripts from interviews, documents, memos, audio files, and so on. Therefore, there needs to be a way to organize and analyze the raw data to answer the research questions and provide a deeper understanding of the phenomenon being studied that is meaningful to the reader. This process often includes countless hours to read, organize, and prepare field notes, transcribe data, code data, categorize data, and identify supporting data. As mentioned, there are many different types of qualitative research designs and ways to analyze and report the narrative data. For more detailed information on qualitative data analysis and reporting, I strongly recommend taking a course in qualitative research or examining textbooks on this topic. However, from my experience advising graduate students, the three most commonly used data collection methods are interviews, focus groups, and observations. Therefore, I will focus my discussion on how to report these types of narrative data according to (a) major themes and patterns, and (b) research questions. I will also discuss how to enhance the validity of the findings.

Major Themes and Patterns

One of the common ways to report findings from narrative data is to organize them around the major themes and patterns. Where do these major themes and patterns come from? Unlike quantitative studies where the researcher has a preset hypothesis that he tests, in qualitative research, you do not start with preset themes and patterns. Instead, the major themes and patterns *emerge* during the data analysis process. For you chefs out there, think of this process as "reducing" sauces where you are producing a thicker, more flavorful, and concentrated sauce. Let us use the example of a case study where a researcher wants to examine business managers' communication styles. She spends six weeks taking field notes as a nonparticipant observing the communication interactions between managers and their employees during weekly two-hour staff meetings. After each meeting, she sits in her car and reflects on what she saw and heard and audio records her thoughts. That evening while cooking dinner, she listens to the audio recording and writes a memo to herself for more clarifying thoughts. At the end of three weeks, the researcher reread and coded the data from her field notes by labeling different topics that seemed important. She puts a code for "interaction" every time there was a communicative interaction between the manager and the employee and noted it in her codebook with a number and definition. Next, she analyzed all the "interaction" codes to see how they were related across participants and meeting times. She categorized or grouped the similar "interaction" codes into larger meaningful chunks with a new label. For example, perhaps there was a clear distinction in managers who used verbal versus nonverbal interactions or collaborative versus coercive interactions. These larger chunks indicate a potential theme or pattern in the findings. Thus, by coding and recoding the data, the research identified the major themes or patterns which emerged from the data. An average of five to six major themes is reasonable. However, be sure the major themes and patterns are related to the research questions and purpose of the study. Remember that it is critical to leave your desires, judgments, and expectations of what you want the data to look like outside of the analysis and reporting process. Although it is inevitable that some personal bias will slip through, you want to minimize this as much as possible.

After the major themes or patterns are identified, each one represents a separate heading and section in the results chapter. Then for each theme, it is critical to paint a "picture" of the findings for the reader by providing a rich and thick description. A **thick description** is an explanation that includes both the behavior and the context in which the behavior was displayed. The concept of thick description was originally derived from the writings of British philosopher Gilbert Ryle and anthropologist Clifford Geertz (Ponterotto & Grieger, 2007). Then sociologist Norman Denzin expanded the definition of thick description in his work:

> Thick description evokes emotionality and self-feelings. It inserts history into experience. It establishes the significance of an experience, or the sequence of events, for the person or persons in question. In thick description, the voices, feelings, actions, and meanings of interacting individuals are heard. (Denzin, 1989, p. 83, as cited in Ponterotto & Grieger, 2007)

As part of the thick description, key pieces of evidence from multiple sources that support the major theme should be included. One major piece of support is in the form of the participants' quotations. This brings the participants' perspective into the study (Creswell, 2012). Keep in mind that you do not want to include everything that was said; instead, quotations should be selected carefully to represent the major theme. This will require you to interpret or infer the participants' true meaning while trying to stay as unbiased as possible. The descriptions could also involve the setting and participants, and the use of visuals in the form of tables or figures should be used to supplement the narrative description.

Here is an example of findings reported around major themes adapted from a former student's master's thesis (Mireles, 2004):

> The informal group discussion provided a wealth of knowledge to the researcher with regards to the elementary students' perceptions around their disabilities. During the discussion, the researcher read from the preselected text. At the end of each subtopic, the following questions were addressed: Can you relate to anything in the passage or can you make a personal connection? The discussions were then left purposely unstructured to allow the students to speak freely and openly but with the guarantee of confidentiality. Four patterns of responses emerged from the discussion group: *feelings about learning, disappointing others, how learning disabilities make you feel*, and *types of learning problems*.
>
> *Feelings about learning*. The students were asked if they could relate to the statement: For some kids, school is not fun because they have trouble succeeding, and they just do not feel good about learning. All six students were able to verbalize that they do not always feel good about learning. Some students gave concrete examples such as Cesar saying, "I don't feel good when I can't do my work in class."
>
> *Disappointing others*. The students were asked to make a personal connection to the provided statement: Some kids feel like they let down their loved ones. All six students stated that they could relate to disappointing their parents, teachers, or themselves. Cesar stated, "I disappoint my parents and teachers every day when I don't do my work and I act out in class." At first Jessica and Charlie were both reluctant to state that they disappointed anyone. Then Jessica said, "I know that my parents don't get mad when I do things wrong. But I know that I disappoint myself because there are things that I can't do. I try and try but I can't. Someone always tells me that I am wrong. This is why I sometimes don't want to come to school." Charlie stated that he knows he disappoints his teachers because, "I just don't get things, especially math. Even when things are explained over and over, I still have trouble."
>
> *How learning disabilities make you feel*. This section provided a lot of opportunity for discussion. When the students heard the word "dumb" in the text, some students verbalized that they felt dumb at times. However, Jessica was also able to express that having a type of learning disability does not mean that you are dumb. She made the personal connection that her father also had a learning disability, and he too had difficulty in school.
>
> *Types of learning problems*. In this section, the students had the opportunity to relate to the difficulties faced by students with learning disabilities. The text discussed difficulties such as memory, concentration, and the ability to make friends, and so on. Kenny, Cesar, and Jessica could all relate to the difficulty with concentration. In response to memory problems, Sam stated, "I just get so frustrated because I don't know my multiplication facts. I try and I try but I just don't know. I am never going to learn them so I just have to add." There were not many students who reported difficulty making friends or maintaining friendships, which appeared to be a sensitive topic.

Research Questions

Another way to organize the narrative findings is around the research questions. Here, the researcher is also reporting the major themes and patterns that emerge from the data. However, in this type of organizational format, the data collection methods (e.g., interview questions, observational strategies) should help you gain an understanding to answer the research question(s) (Maxwell, 2013). For example, a researcher wants to explore the leadership styles of school administrators at high-performing schools. The research questions are,

1. What are the leadership characteristics of administrators at high-performing schools?
2. How do these administrators overcome barriers to success?

She conducts one-hour interviews with four administrators from high-performing schools. There are four main questions that she asks at each interview:

1. What is your leadership style?
2. What makes you an effective leader?
3. What are the barriers that you face as a school administrator?
4. How do you get faculty and staff to support your leadership style?

In this study, the four interview questions are designed to help answer the research questions; interview questions 1 and 2 help to answer the first research question, and interview questions 3 and 4 help to answer the second research question. These four questions are just a start; other questions or follow-up questions may be necessary as the interview proceeds. If possible, you should always pilot test your data collection methods with a similar sample group and setting to see if any revisions are necessary.

After conducting the interviews, the researcher must still follow a process for data analysis. First, she reads through and transcribes all of the interview data. Next, she codes the data from the transcripts by labeling different topics. Then, she categorizes or groups the codes into larger meaningful chunks with a new label. However, the key difference with this organizational format is that she pulls out major themes only from interview questions 1 and 2 to answer the first research question. For example, perhaps there was a clear pattern that leaders believed collaboration was a key component of their effective leadership style. Then she pulls out major themes from interview questions 3 and 4 to answer the second research question. After the major themes are identified, then rich, thick descriptions with supporting evidence and quotations are reported.

Here is an example of findings reported around research questions adapted from a former student's master's thesis (Kendall, 2006):

> An analysis of the data yielded from the student and staff questionnaires revealed findings within the areas of the research questions. Student and staff participants' responses to the questionnaires were grouped to correspond to the research questions and then categorized for major themes or patterns.
>
> Research question 1 asked what factors of communication (whether verbal or nonverbal) triggered negative behavioral outbursts or promoted positive and effective communication in classrooms serving students with emotional disturbances (ED) and learning disabilities (LD). The data revealed that the verbal factors of communication that triggered students' behavioral outbursts were yelling, especially once the student was already upset. Other factors included students feeling like they were not being understood or listened to, not getting help with their assignments, and negative peer interactions in the classroom. The nonverbal factors of communication that triggered students' negative behavioral outbursts were slamming books down and making angry faces.
>
> The data revealed various verbal factors that promoted positive behavior and effective communication in classrooms serving students with ED and LD. Some of these factors were taking the time to discuss classroom issues with the students in a calm voice, giving the students some extra chances, the implementation of classroom reward systems, explaining the lessons thoroughly when needed, and positive peer interactions. The nonverbal factors that promoted positive behavior and effective communication were allowing the students space when their behaviors were escalating.
>
> The questions on the students' questionnaires that corresponded to this research question were questions 1, 4, 7, 8, and 10.
>
> Question 1 asked what the staff should or should not do to help when the students were having a particularly rough day. Most students responded that yelling would only escalate their behavior, and that the teachers should either speak to them about whatever the problem was in a calm voice or give them some extra chances. For example, one male Caucasian student in the 11th grade said, "They should tell me to cool down or give me a break outside. They shouldn't get on my back when I am mad." A few students felt the need to be left alone when they were having rough days. For example, one female African American student in the 12th grade said, "[If I am having a bad day] they shouldn't do anything because I will still get mad."
>
> Question 4 asked for the reasons behind the students' best and worst behaviors. Most students attributed their best behavior to factors outside of the classroom such as having a good night sleep, a good breakfast, good weather, or positive experiences with friends prior to class. For example, one male Caucasian student in the 11th grade said, "I woke up on the right side of the bed and played with my brother. That made me happy." Other participants reported that on days when their behaviors were at their best, they were connected to factors inside the classroom such as classroom rewards, having lessons explained to them well, or positive student interactions. For example, one male African American student in the ninth grade said, "[The day my behavior was at its best] was the day I hit level 5."
>
> With regard to negative behavior, a few students attributed the behavior to factors outside of the classroom such as showing up in a bad mood. Most attributed their negative behavior to factors inside the classroom such as the teacher yelling at them, not being understood or listened to, not getting help with assignments, or negative peer interactions. For example, one male Caucasian student in the 11th grade responded, "When my teacher always yells and gets in my face and I get mad and punch the walls." Another male Latino student in the 11th grade reported, "They don't even listen to me, and they act like I wasn't even there. That's why I had the worst behavior in Mrs. C.'s class because I don't get any help with my work."

Validity of Findings

Regardless of the format that is chosen to report findings, an important component of reporting qualitative data is to ensure their validity. Validity in this context is about the quality of the findings, which is different than the validity with regard to quantitative measures that was mentioned in Chapter 7. For qualitative studies, validity refers to the accuracy and credibility of the findings (Creswell, 2009). In other words, are the findings plausible and trustworthy? You increase the validity of a qualitative study by applying strategies to reduce factors that threaten the credibility of your conclusions. For example, personal bias (subjectivity) is a validity threat because your own experiences, assumptions, goals, and beliefs will influence how you analyze, interpret, and present the data. There are many strategies available to increase validity, and you should apply those that are specific to the validity threats in your study. Some recommended strategies include being upfront and honest about your own personal biases, having extensive time and opportunities for data collection, checking back with participants for accuracy (i.e., member checking), and providing rich and thick descriptions (Creswell, 2013). A commonly used method to increase validity is triangulation. Triangulation is "the combination of methodologies in the study of the same phenomenon" (Denzin, 1978, p. 291, as cited by Onwuegbuzie, 2002). **Data triangulation** is one form of triangulation where multiple methods of data collection are used to study one phenomenon. The different methods act as a "check" on the others to support a single conclusion or provide new information from different angles (Maxwell, 2013). For example, a researcher could use multiple data collection methods such as observations, interviews, and written documents. The point is not to combine the data but rather to find the intersections, conflicts, or connections among them. In doing so, the researcher is able to confirm, dispute, or corroborate findings between data sources and have a holistic picture of the phenomenon.

Summary

Chapter Four is a significant chapter in the thesis because it reports the major results and findings of the study to the reader. Chapter Four may also be one of the most satisfying chapters to write because after all the months of data collection and data analysis, you finally get to share all that was discovered. Whether you are reporting quantitative or qualitative data, it is critical to be as detailed and comprehensive in your descriptions as possible. This will enhance the validity, quality, generalizability, or transferability of the results. In this chapter, you are also laying the foundation for the final discussion and conclusions, which are the focus of Chapter Five, Discussion, of your thesis. This will be the topic of the next chapter of the book. Here is a summary of the most critical points from Chapter 8:

- Depending on the research questions and design, the presentation of the results can be in narrative, numerical, or tabular/graphic format.
- Before reporting the results, make sure that all your data have been organized and analyzed.
- In descriptive statistics, measures of central tendency, such as the mean, median, and mode, tell you the "average" score in a distribution.
- Measures of variability, such as the range or standard deviation, tell you how close or spread apart (i.e., dispersed) the scores are in a distribution.
- Inferential statistics use sample group data to make assumptions about the general population.
- In experimental studies, tests of significance are used to determine if observed mean differences between groups or conditions represent a real difference or were due to chance.
- The null hypothesis, H_0, represents the "chance" theory, meaning any observed differences are due to chance, and the treatment has no significant effect on the dependent variable.

- One of the ways to report findings from narrative data is to organize them around the major themes and patterns that *emerge* during data analysis.
- Another way to organize the narrative findings is around the research questions where items from the data collection instruments help to answer the research question(s).
- There are several ways to increase the validity of a qualitative study, such as being upfront and honest about your own personal biases, having extensive time and opportunities for data collection, member checking, providing thick descriptions, and data triangulation.

Resources

Common Obstacles and Practical Solutions

1. A common obstacle that students face in writing Chapter Four is being overwhelmed with the amount of results to report. Words that come to mind are, "How do I make sense of all these data?" If you have quantitative data, the best way to overcome this obstacle is to report the results in chunks. First, look at the total data set. Then, examine the data to see which method would make the most sense to report in terms of organization (usually by each measurement instrument). Next, decide whether you should report the data with descriptive statistics, inferential statistics, or with visual representation such as figures and graphs. Definitely get help from your chairperson if you do not understand the results from the data analysis. Then, report the results from one measurement instrument and have your chairperson review it before you go on to do the rest.
2. Another common obstacle faced by students is finding the major themes and patterns in the qualitative data. Although the major themes and patterns do emerge from the data, sometimes it is not obvious as to what they are. Words that come to mind are, "How do I tie all these together?" After coding the data for specific topics, you need to step back and look at the data from a broader perspective. Sometimes, you have to recode the data into larger categories. Using multiple highlighting colors to code or physically cutting and grouping "like data" together may also help to find the themes and patterns. You may also want to invest in a computer software program to help you with the data analysis. One thing to always keep in mind is the purpose of the study and the research questions. If you use these as your guiding principles, this will help to make sure you do not get lost in all the trees.

Reflection/Discussion Questions

When you report your data in Chapter Four, it is important to understand the differences in how to organize the results (depending on the type of data collected) so that it is meaningful to the reader. This is also important for replicability purposes in case someone is interested in confirming or corroborating the results. The following reflection/discussion questions will help guide you through the reporting process.

1. What is the difference between descriptive and inferential statistics? Give examples of measures of central tendency and variability. Give examples of when you would use an independent-samples *t* test versus a paired-samples *t* test. Then, pick one specific measurement instrument that you used in data collection and discuss what type of statistics you would use to report the results.

2. What is the definition of a "thick description"? Why is having a thick description important in reporting qualitative data? Give examples of what information you could provide in the thick description to increase validity.

Try It Exercises

The following exercises are designed to help you write Chapter Four. In Activity One, you will calculate the descriptive statistics for a given data set. In Activity Two, you will identify the inferential statistics that will be used to report the data. In Activity Three, you will report the findings from one measurement instrument that you used to collect data.

1. Activity One: For this activity, use your knowledge of descriptive statistics.

A researcher conducted a study on the effects of an online course to teach nursing students how to provide proper drug dosage calculations for their critical care patients. The students were randomly assigned into two groups: online course (Group A) or traditional course with instructor (Group B). The following data set represents their posttest scores on a drug dosage calculation test. With a partner, identify the descriptive statistics for each set of scores.

Group A:	3	4	4	9	1	15
Group B:	3	17	12	4	3	3

- What are the mode scores for Groups A and B?
- What are the median scores for Groups A and B?
- What are the mean scores for Groups A and B?
- What are the range scores for Groups A and B?
- Which group do you think has a larger standard deviation?
- Report the findings for each group in APA format.

2. Activity Two: For this activity, use your knowledge of inferential statistics.

Now the researcher wants to know if there was a statistically significant mean difference between the two groups. With a partner, identify the inferential statistics using the scores from Activity One.

- What test of significance should the researcher use to analyze the mean difference?
- Write a null hypothesis for the research study.
- What should the researcher set the significance (alpha) level at, and what does this mean?
- What would be the degrees of freedom for the test?
- If the *t* score is 4.52 and the probability value is .03, should the researcher reject or retain the null hypothesis?
- What is the researcher's final conclusion about the online course?
- Report the findings in APA format.

3. Activity Three: For this activity, focus on the data from one measurement instrument that was used during the study.

- Pick one measurement instrument that you used to collect data such as a test, survey, interview questions, and so on.
- Decide the best way to report the data. If you collected numerical data, decide whether to report descriptive or inferential statistics. If you collected nonnumerical data, decide whether you want to report the major themes/patterns from the entire data set or want to connect specific items to corresponding research questions.
- Prepare a draft report of the results and have a partner or your chairperson review it before proceeding with the other data.

Key Terms

- data triangulation
- degrees of freedom
- descriptive statistics
- independent-samples *t* test
- inferential statistics
- mean
- measure of central tendency
- measure of variability
- median
- mode

- null hypothesis
- paired-samples *t* test
- range
- standard deviation
- *t* test
- thick description
- Type I error

Suggested Readings

- Angrosino, M. V. (2005). Recontextualizing observation. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 729–745). Thousand Oaks, CA: Sage.
- Curran-Everett, D. (2009). Explorations in statistics: Hypothesis tests and P values. *Advances in Physiology Education, 33*, 81–86.
- Eisenhauer, J. G. (2008). Degrees of freedom. *Teaching Statistics, 30*(3), 75–78.
- Elliot, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology, 38*, 215–229.
- Fischer, C. T. (1999). Designing qualitative research reports for publication. In M. Kopala & L. A. Suzuki (Eds.), *Using qualitative methods in psychology* (pp. 105–119). Thousand Oaks, CA: Sage.
- Fontana, A., & Frey, J. H. (2005). The interview: From neutral stance to political involvement. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 695–728). Thousand Oaks, CA: Sage.
- Glesne, C. (2011). *Becoming qualitative researchers: An introduction* (4th ed.). Boston: Pearson Education.
- Krause, S. (2010). Monstrous methods for teaching central tendency concepts. *Teaching Statistics, 32*(1), 21–23.
- Little, L. (2010). *Reporting results of common statistical tests in APA format.* Available online: http://web.psych.washington.edu/writingcenter/writingguides/pdf/stats.pdf
- Morrow, S. L. (2005). Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counseling Psychology, 52*, 250–260.
- Ponterotto, J. G., & Grieger, I. (2007). Effectively communicating qualitative research. *The Counseling Psychologist, 35*(3), 404–430.
- Sands, R. G., & Roer-Strier, D. (2006). Using data triangulation of mother and daughter interviews to enhance research about families. *Qualitative Social Work, 5*(2), 237–260. doi:10.1177/1473325006064260

Web Links

- Psychological Statistics http://www.uwsp.edu/PSYCH/stat/5/CT-Var.htm#II
- QSR International (NVivo) http://www.qsrinternational.com/products_nvivo.aspx

- Reporting Statistics in APA Style: A Short Guide to Handling Numbers and Statistics in APA Format http://my.ilstu.edu/~mshesso/apa_stats.htm
- Reporting Statistics in APA Style http://www.ilstu.edu/~jhkahn/apastats.html
- Researchware (HyperRESEARCH, HyperTRANSCRIBE) http://www.researchware.com/
- Using SPSS to Understand Research and Data Analysis http://wwwstage.valpo.edu/other/dabook/home.htm