RESEARCH IN COMMUNICATION SCIENCES AND DISORDERS

METHODS FOR SYSTEMATIC INQUIRY

Third Edition

RESEARCH IN COMMUNICATION SCIENCES AND DISORDERS

METHODS FOR SYSTEMATIC INQUIRY

Third Edition

Lauren K. Nelson, PhD, CCC-SLP

PLURAL PUBLISHING INC.

PLURAL PUBLISHING
INC.

5521 Ruffin Road
San Diego, CA 92123

e-mail: info@pluralpublishing.com
Website: http://www.pluralpublishing.com

Typeset in 10.5/13 Garamond Book by Flanagan's Publishing Services, Inc.
Printed in the United States of America by McNaughton & Gunn

Library of Congress Cataloging-in-Publication Data:

Names: Nelson, Lauren, author.
Title: Research in communication sciences and disorders : methods for systematic inquiry / Lauren K. Nelson.
Description: Third edition. | San Diego, CA : Plural Publishing, [2017] | Includes bibliographical references and index.
Identifiers: LCCN 2016027632| ISBN 9781597567268 (alk. paper) | ISBN 1597567264 (alk. paper)
Subjects: | MESH: Communication Disorders | Biomedical Research—methods | Research Design
Classification: LCC RC423 | NLM WL 340.2 | DDC 362.196/8550072—dc23
LC record available at https://lccn.loc.gov/2016027632

CONTENTS

Types of Research

Taking some time to peruse published research in audiology and speech-language pathology, such as that found in our professional journals, reveals many forms of research. Generally, research studies share at least one similarity: a question that needs an answer or problem that needs a solution. How researchers formulate their questions or how they plan and conduct their studies, however, can be quite different. In this section we consider some of the terminology researchers use to characterize these differences.

Most of the time when professionals in the fields of audiology and speech-language pathology use the term *research*, they are referring to *empirical* research. Empirical research involves the collection of new information or data through observation and measurement of behavior and/or physical properties (Trochim, Donnelly, & Arora, 2016). Review of recent issues of professional journals in communication sciences and disorders reveals several ways that human behavior was observed and measured, such as speech samples (Ertmer & Jung, 2012), survey responses (Teten, DeVeney, & Friehe, 2016), listener ratings (Isetti, Xuereb, & Eadie, 2014), questionnaire responses (Umansky, Jeffe, & Lieu, 2011), and test scores (van Kleeck, Vander Woude, & Hammett, 2006), as well as several ways of measuring physical properties: speaker sound pressure levels (Spielman, Ramig, Mahler, Halpern, & Gavin, 2007), tongue strength and endurance (Stierwalt & Youmans, 2007), otoacoustic emissions (Spankovich et al., 2011) and electromyographic waveform displays (Walsh & Smith, 2013).

Researchers conducting nonempirical investigations make use of existing information instead of gathering new data. Nonempirical research ranges from loosely structured term papers and literature reviews to carefully constructed theoretical analyses or systematic reviews of a body of research.

Another way to characterize different forms of research is the distinction between *qualitative* and *quantitative* research. Qualitative research and quantitative research differs with regard to the way questions or problems are formulated and investigated. A commonly identified difference, however, is in the type of information or data a researcher gathers. Qualitative research data often include verbal information. This might take the form of highly detailed descriptions of a person's behavior or perhaps direct quotes of a person's statements. Quantitative research, as you might expect, relates to numerical information such as frequency counts and measures of size or other physical properties. Sometimes researchers gather both types of data and report both numerical and verbal information.

Within the category of quantitative research we often make a distinction between studies that are *experimental* and those that are *nonexperimental*. In experimental research, researchers identify one or more factors that they will manipulate or control during the experiment. For example, a researcher might compare different approaches for improving a person's communication abilities and could manipulate how much or what type of approach participants experience. The researcher manipulates or controls the conditions so that some participants have a different experience during the experiment than others. According to Patten (2013), a *true experiment* meets two criteria. The first is the researcher's creation of different conditions or experiences by manipulating one or more factors during the experiment, and the second is that the conditions participants experience are determined randomly. A true experiment

has random assignment of the participants to different experimental groups. Experimental research that lacks random assignment to groups is sometimes referred to as *quasi-experimental* research. Generally speaking, a study that meets both standards, experimental manipulation and random assignment, provides stronger evidence than a quasi-experimental study.

One of the most common kinds of experiments is one in which a researcher compares the performance of two groups, each experiencing a different experimental manipulation or treatment. In audiology and speech-language pathology such comparisons might involve traditional treatment as compared to some new treatment approach. As noted, when the participants are divided at random into groups, the study is considered a true experiment. Sometimes researchers find it impractical or impossible to assign their participants randomly, however. Perhaps the researchers want to compare two different classroom-based interventions. Children in school settings are seldom assigned to their classrooms in a random manner. Therefore, if researchers decide to conduct the experiment with two existing classrooms, they are conducting quasi-experimental research rather than a true experiment.

In contrast with experimental research, nonexperimental research includes a wide variety of studies in which the researcher investigates existing conditions. Some forms of nonexperimental research are descriptive in nature. Studies that provide information about the typical communication behaviors of persons of various ages fall into this category. Such studies might include measures based on speech and language samples, measures of physical properties of speech, such as fundamental frequency or intensity, as well as psychoacoustic responses to speech. Other examples of nonexperimental research include case studies, surveys, studies of relationships or correlations between measures, as well as comparison or case-control studies (Patten, 2013). In comparison or case-control studies, researchers include groups of persons with pre-existing differences, rather than create differences via an experimental manipulation. Some examples include comparisons of 3-year-olds and 5-year-olds, persons with a particular type of hearing loss and persons with normal hearing, adults with functional voice disorders and those with normal voices, or children with specific language impairment and those with typical language.

Another important distinction for research in communication sciences and disorders is the difference between *group* and *single subject* research. This difference is not associated with the number of research participants in a literal way. That is, one might encounter a small group study with just five participants in each group or one might encounter single subject research with several participants. Nor is single subject research the equivalent of a case study. Case studies involve nonexperimental, descriptive research, whereas single subject research is experimental in nature. The most important differences between group and single subject research concern how participants are treated during the study and how their data are reported. In single subject research, a participant experiences both the experimental and control conditions, and results for each participant are reported separately. When experimental and control conditions are compared in group research, usually the participants in one group experience the experimental condition and the participants in another group experience the control condition. The results from group research are aggregated and reported for each of the comparison groups and not for individual participants.

APPENDIX 2–1

Research Scenario

Responsible Conduct of Research Scenario

Please note that the following case description is a work of fiction. It is not intended to represent any actual individuals or events.

Dr. P. T. Smith has worked as an audiologist in an outpatient clinic at a small hospital since graduating with an AuD degree approximately four years ago. As a graduate student, PT embraced the concept of clinician-investigator and completed an empirical research project under the direction of audiology faculty member, Dr. R. Star. PT particularly enjoyed Dr. Star's mentorship and individual attention as they worked together on the research project. Knowing PT's interest in completing a research project, Dr. Star had recruited PT to work on an idea that was already partially developed. Dr. Star had the notion of developing a follow-up program for new hearing aid users based on adult learning theory. PT liked the idea right away and could appreciate its clinical relevance. Dr. Star's idea was relatively undefined, however, and PT had put in considerable time studying the literature on adult learning: developing a script for the follow-up training, planning a set of short activities to orient new patients to the features of their devices, and developing some listening activities that simulated the experience of listening in different conditions. PT and Dr. Star obtained permission to run their study from their University's Institutional Review Board. The design of the study involved randomly assigning persons who were new hearing aid users to either PT's new follow-up training or the traditional follow up that had been used in the campus clinic for many years. The participants completed a user satisfaction questionnaire one month after receiving their hearing aids and six months later. Ultimately, because of the time spent in developing the training protocol, PT only had time to run 10 individuals through the study, five who completed PT's training protocol and five in the control group. PT's only disappointment with the research was that they did not find any significant differences in user satisfaction. Even though the mean scores for PT's experimental follow-up procedures were higher than for the control group, these differences were not very strong. Dr. Star still praised PT's work and stated that the study would be a very good pilot study for future work on the topic.

About a year after PT graduated, R. Star took a job at a larger university known for its strong research programs. PT kept track of R. Star's work through the audiology literature. One day when perusing one of the audiology journals, PT was surprised to see an article by R. Star and a new doctoral student. The article was on the same topic as PT's graduate research project. In reading the article, PT noted one sentence acknowledging "preliminary work on this topic in an unpublished research paper" (Smith, 2008). When PT read the full article, however, it seemed as though the methods in this new paper were identical to the ones PT had developed for the smaller study four years previously. PT was disappointed that Dr. Star had not acknowledged this contribu-

tion in the methods section. PT would have enjoyed publication credit and wondered if the pilot study and work on the methods warranted inclusion as an author. In a sense, PT felt betrayed. Shouldn't Dr. Star have at least acknowledged PT's role in developing the experimental training protocol? Rightly or wrongly, PT felt some sense of ownership over the experimental protocol.

Discussion Questions

1. What are the issues in this case that relate to research ethics?
2. In your opinion, how should Dr. Star have acknowledged PT's work on the experimental training protocol? Was the brief mention about preliminary work sufficient? Be prepared to explain your answer.
3. Would your answer to question 2 change if Dr. Star used the data from PT's study and simply collected data from more participants for the published work?
4. How could researchers such as PT and Dr. Star avoid this type of conflict in the future?

existing groups to identify differences by presence of a disorder, age, socioeconomic status, and so forth; and causal-comparative and cohort studies that examine the impact of possible causal factors over time (Patten, 2013). In the next sections, we cover several common nonexperimental designs and discuss their use in the field of communication sciences and disorders.

Survey Research

You might have personal experience with *survey research*. Perhaps a researcher has asked you to complete a paper and pencil or telephone survey on some topic, such as your opinion on a certain product or on election issues. Survey research generally involves obtaining participants' responses to a series of questions, either through a written questionnaire or an interview (Trochim et al., 2016). Researchers might consider using a survey when they want to collect data that reflect opinions or reports of individual experiences and when they want to collect information from a relatively large number of participants (Writing@CSU | The Writing Studio, 1993–2016). When designing a survey, a researcher needs to decide on several components, such as those listed below:

1. Survey participants
2. Content of questions
3. Types of questions
4. Sequence of questions
5. Survey procedure (e.g., written or interview)

The subject matter of a survey is the most important factor to consider when deciding who should complete the survey. For example, if the survey focuses on consumer satisfaction with speech, language, or hearing services, the survey participants should be the persons who received those services rather than the audiologists and speech-language pathologists who provided the services. Sometimes researchers have choices regarding the most appropriate participants. For example, if a survey focuses on children's actions and attitudes toward peers with communication disorders, the survey participants could be the children or perhaps their teachers. Additionally, the researchers might decide to compare responses from different groups of participants. They might survey different age groups, different professions (audiologists, speech-language pathologists, speech-language-hearing scientists), persons in different geographic locations, and so forth.

The content of survey questions relates closely to the validity of the survey or the extent to which it covers the material professionals in the field would expect. Furthermore, survey researchers also need to consider if they are asking for information their respondents know, if the wording of the questions elicits appropriately specific information, and if the terminology in the questions is familiar to respondents (Trochim et al., 2016). A survey can include several types of questions including yes/no, categorical response, rating scale, semantic differential, cumulative response, and open-ended formats (Trochim et al., 2016). Table 5–1 includes examples of each type of question.

In addition to making decisions about the kinds of questions to use, survey researchers also need to make decisions about the sequence of questions, as well as how to administer the questions to participants. Trochim et al. (2016) suggested that surveys should start with straightforward questions and present probing or difficult questions toward the end. In deciding

Table 5–1. Examples of Five Different Types of Survey Questions

Type of Question	*Example*
Yes/No	1. Are you currently employed as an audiologists or speech-language pathologists? (Circle one) Yes No
Categorical Response	2. What is your current class standing? (Place an X beside one) ____ Freshman ____ Sophomore ____ Junior ____ Senior ____ Graduate
Rating Scale	3. Evidence-based practice will improve the quality of patient care in audiology and speech-language pathology. Strongly Agree 1 2 Neutral 3 4 Strongly Disagree 5
Cumulative Response	4. In the past month, I provided speech, language, or hearing services to adults or children with: (Place an X beside all that apply.) ____ articulation/phonological disorders ____ auditory processing disorders ____ conductive hearing loss ____ fluency disorders ____ language disorders ____ sensorineural hearing loss ____ voice disorders ____ other (_____)
Open-Ended	5. In your opinion, what are the three most important issues facing the field of communication sciences and disorders?

whether to use a written survey or interview, researchers should consider the advantages and disadvantages of each approach. Advantages of written surveys include the possibility of displaying graphic or pictorial content, greater privacy for respondents, relatively low cost, and ability to recruit participants from a wider geographic area. Advantages of interviews include being able to explain the survey and answer participant questions, modify questions and ask follow-up questions, and include respondents who do not read or write (Trochim et al., 2016).

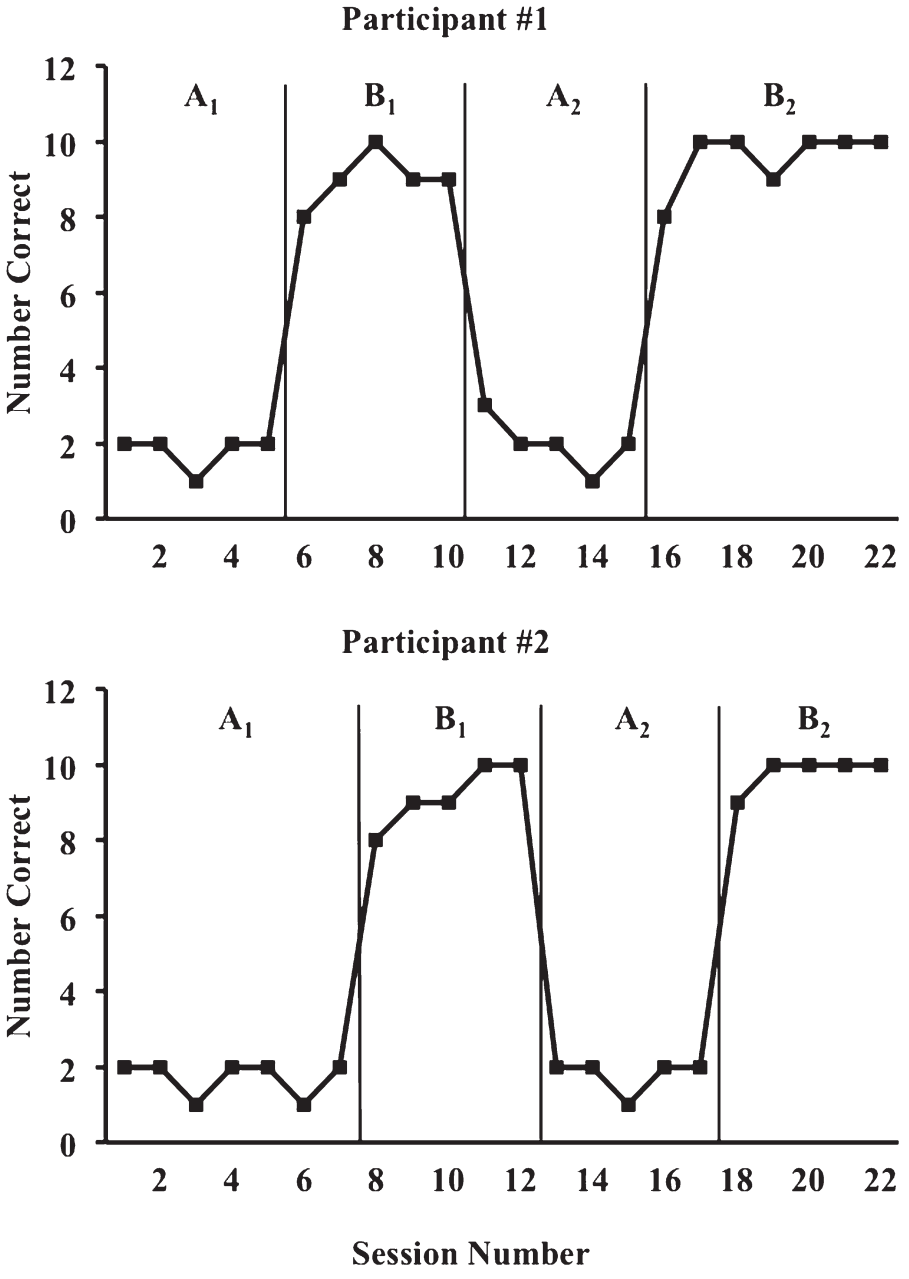

Figure 6–5. Illustration of an A_1–B_1–A_2–B_2 multiple baseline across participants single-subject design.

develop a working definition of the behaviors to assure reliable measurement (Horner et al., 2005; Wendt & Miller, 2012). Fourth, researchers need a detailed description of the treatment or independent variable so that it can be administered consistently and

replicated across experimenters. Finally, researchers should arrange for independent measurement of relevant behaviors by individuals who are unfamiliar with the participants and phases of the study, that is, blinded to the baseline versus experimental treatment conditions (Reichow et al., 2008; Tate et al., 2013).

Single-subject research designs are well suited to studying an individual's response to intervention (Leary, 2011; Lillie et al., 2011). If individual participants respond in unique ways to the intervention, the research is still valuable because findings are not aggregated across participants. In group research, individual variability in response to intervention tends to be obscured in the averaged data. These designs allow researchers to compare behaviors observed during baseline phases with those observed during intervention to determine if there is a functional, cause and effect relationship between an intervention and observed changes in behavior (Byiers et al., 2012). Because single-subject designs provide a researcher with ways to establish experimental control, they provide stronger evidence than nonexperimental case studies. Single-subject designs have a long history of use in the field of communication sciences and disorders in research with both children and adults, and Table 6–6 provides a brief summary of several recent studies that employed single-subject experimental designs.

Experimental Designs and Levels of Evidence

Previously we learned that the term *evidence-based practice* (EBP) refers to an approach in which clinicians use the best available research to guide their decisions about how to evaluate and treat persons with communication disorders. Some of the steps in EBP include identifying a clinical question, searching the professional literature, and reading and evaluating research reports. In Chapters 3 and 4, we covered developing a clinical question and conducting a literature search as the initial steps in completing evidence-based practice research. The next step after completing the literature search is to select the most relevant articles to read and evaluate. An important consideration in evaluating research is to determine the *level of evidence* a study provides, and knowledge about research design is a key factor in making this judgment. As noted previously in this chapter, the strongest designs for establishing cause and effect relationships are true experimental designs: those with random assignment of participants to at least two groups, a treatment and a control group. True experimental designs, sometimes called randomized clinical trials, provide the strongest kind of evidence for intervention effectiveness and include studies that compare treatment and no-treatment conditions, as well as those that compare two or more different treatment approaches. Other experimental designs provide useful information, but the strength of evidence for a cause and effect relationship is weaker.

In addition to level of evidence, audiologists and speech-language pathologists consider the depth of evidence that supports the effectiveness of a particular intervention approach. For example, a well-designed, randomized clinical trial is considered a strong type of evidence, and several well-designed, randomized clinical trials that yielded similar results would be even stronger. If the studies yielded conflicting results, then the evidence in support of a particular intervention approach is undermined.

Sometimes your literature search might yield a systematic review or meta-analysis.

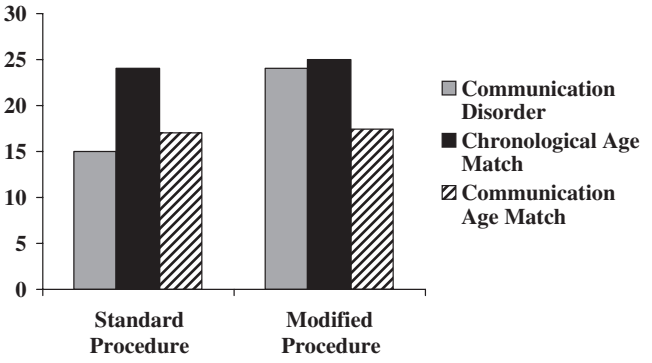

Figure 8–3. Illustration of the use of a column graph to show group means by experimental condition.

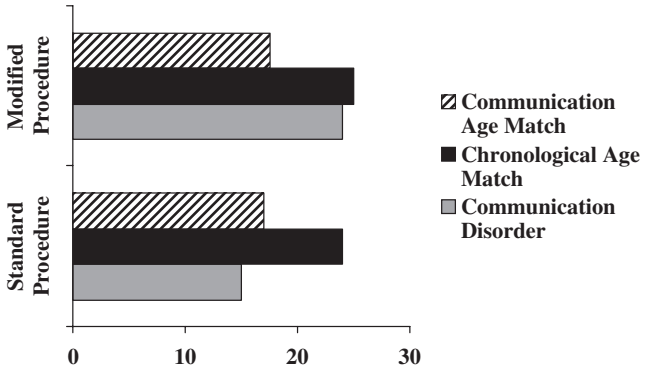

Figure 8–4. Illustration of the use of a bar graph to show group means by experimental condition.

and line graphs are similar because all three types are useful for illustrating different values for frequencies, counts, percentages, and averages. Depending on the nature of your data, these values could be associated with different groups, different tasks, different behaviors, changes over time, and so forth. Line graphs might be particularly suitable for depicting several values in a series or for depicting a special kind of nonlinear relationship called an interaction. An interaction occurs when two or more groups respond in unique ways to the experimental manipulations. The example line graph in Figure 8–5 illustrates an interaction in a simple 2 × 2 design, that is, a design with two different groups (e.g., persons with a hearing impairment or those with normal hearing) and two levels of the experimental manipulation (e.g., standard or modified test procedure). This example shows the use of a line graph to depict group means as well as variability within each group, as shown by the vertical lines extending from each

samples are either two measures on the same participants, such as measures at two different times, or measures from samples of matched participants. Sometimes researchers create participant pairs by matching them on some pretest measure. After the matching the researchers randomly assign the participants to one of two treatment groups. With either two measures on the same participants or matched participants, the appropriate *t*-test is a *paired t-test*. With independent samples, the appropriate *t*-test is an *independent t-test*. One way this choice affects your analysis is when you determine degrees of freedom. The degrees of freedom for a paired *t*-test are the number of pairs minus one ($n - 1$). If you had 25 matched participants, the degrees of freedom for your analysis would be 24. For an independent *t*-test you have two options for determining degrees of freedom. The most common when you have equal numbers of participants in each group is called the *pooled variance t-test*. You add the number of participants in each group and subtract 2 ($n_1 + n_2 - 2$). If you had 25 participants in each group the degrees of freedom would be 25 + 25 – 2 or 48. When you have two independent samples, and the variances of the two groups are different, using the pooled variance *t*-test is inappropriate. Determining the degrees of freedom for groups with different variances is beyond the scope of our discussion. You might consult other sources if confronted with this situation (StatSoft, 2013).

Let's return to the data in Table 9–4. These data represent independent samples, so we are going to analyze the mean differences using an independent *t*-test. First, we are going to report the results for the larger samples ($n = 25$). Pyrczak (2010) recommends always reporting the values of your means and standard deviations before reporting the results of a *t*-test, so this information is included in Table 9–4. The results from the analysis of the more variable groups revealed a significant difference between the first and second groups, $t(48) = -2.69$, $p < 0.01$. Repeating the analysis for the data in columns 5 and 6 revealed a significant difference between the fifth and sixth groups as well, $t(48) = -5.33$, $p < 0.0001$. Thus, with samples of 25 participants the differences between groups were significant for both the more variable and less variable samples.

Let's repeat our analyses with the smaller samples, starting with the more variable groups in columns 3 and 4. In this case our degrees of freedom will be 10 + 10 – 2 or 18. The results from this analysis revealed the mean difference between the third and fourth groups was not significant at the 0.05 level, $t(18) = -1.82$, $p = 0.09$. Finally, let's complete our example by analyzing the data in columns 7 and 8. The results from the analysis of the less variable groups revealed a significant difference between the seventh and eighth groups, $t(18) = -2.86$, $p = 0.01$.[3] The *t*-test results reported here follow the guidelines in the APA (2010) publication manual by including the statistical symbol, degrees of freedom, statistical value, and probability.

Confidence Intervals

Although difference tests, such as the *t* test are common in published articles, recently investigators have been encouraged to consider alternative procedures such as reporting confidence intervals (Cumming, 2012;

[3]Many dedicated statistical software packages report the actual probability of error, and the *p* values reported by DataDesk® 6.3 are included here.

Kline, 2004). In keeping with this suggestion, let's calculate and display the confidence intervals for the data in Table 9–4. A *confidence interval* is a way of estimating the margin of error associated with your sample. Remember that researchers study a sample in order to make inferences about the entire population from that sample. However, a sample may or may not represent the population well. If you could obtain 100 different samples from a population and calculate a mean from each sample, some of those samples would provide very close estimates of the population mean, but other samples would provide estimates that were somewhat distant from the population mean. This is the notion of margin of error. Researchers calculate a confidence interval to establish a range that has a high probability of capturing the true population mean.

To calculate a confidence interval, we need four numbers: the number of participants (*n*), mean, standard deviation (*SD*), and a critical value for the *t* statistic at a particular level of confidence, for example, .05 or .01. Researchers obtain the first three numbers from their data and look up the final number in a statistical table. The formula for calculating margin of error (*MOE*) and the 95% confidence interval was first described in Chapter 8. The calculation involved determining the standard error (*SE*) and then using that value to determine *MOE* and the 95% confidence interval.[4] Recall that the *SE* was the *SD* divided by the square root of *N*, and that *MOE* was $t_{.95}$* SE,* where $t_{.95}$ was obtained from a table for the *t* statistic, and finally, the confidence interval was the mean ± the *MOE* (Cumming, 2012). Means and standard deviations for each of the groups are already included in Table 9–4. To complete our calculation, we need the *t* for the 95% confidence level which is 2.065 for our larger groups (for *n* – 1 or 24 degrees of freedom) and 2.262 for our smaller groups (for *n* – 1 or 9 degrees of freedom). The confidence intervals (CI) for each group are reported below and plotted in Figure 9–8.[5] Our population means would probably fall within these ranges 95% of the time (i.e., 95% confidence interval).

More variable *n* = 25, 1st group (A1)
CI is 44.54 to 54.66

More variable *n* = 25, 2nd group (A2)
CI is 54.35 to 64.85

More variable *n* = 10, 1st group (B1)
CI is 41.39 to 56.01

More variable *n* = 10, 2nd group (B2)
CI is 51.29 to 64.59

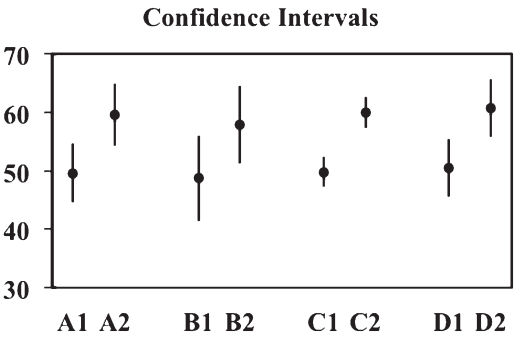

Figure 9–8. A plot showing the confidence intervals for the mean ± the margin of error for each of the groups compared in Table 9–4.

[4]As with effect size, tools for calculating a confidence interval are available via the Internet, for example, the *EasyCalculation.com* website (https://www.easycalculation.com/statistics/data-analysis.php)

[5]Although you could compute these confidence intervals manually with a calculator, these examples were generated using *Exploratory Software for Confidence Intervals (ESCI),* free software that accompanies the Cumming (2012) textbook.

factors to consider in evaluating a study are the breadth of evidence and the source of the evidence. Recall that the highest level of evidence is a meta-analysis or systematic review and that you need several studies on the same topic to prepare this type of document. When evaluating the body of evidence for a particular treatment, you need to consider how many studies are available, whether these studies provide conflicting or converging evidence (ASHA, 2004), and who conducted the study and wrote the research report. A body of evidence from multiple researchers, persons other than those who first developed an approach, and multiple treatment sites would be much stronger than evidence from a single research group (Nail-Chiwetalu & Ratner, 2006; Ratner, 2006).

The various criteria for appraising clinical research are summarized in Table 10–1. Researchers might use criteria such as those in the table when evaluating studies for evidence based practice or when writing a review of literature for an original research report. You also might consider other sources like the critical appraisal worksheets provided by the Centre for Evidence-Based Medicine (CEBM, 2016), the PEDro-Scale (Speech Pathology Database for Best Interventions and Treatment Efficacy [SpeechBITE], n.d.), or the criteria used in recent systematic reviews and meta-analyses (e.g., Gerber, Brice, Capone, Fujiki, & Timler, 2012; McCauley, Strand, Lof, Schooling, & Frymark, 2009). If the research article is a single-subject design study, many of the criteria in Table 10–1 are still applicable, such as the type of design, blinding, long-term posttest, detailed description of participants, evidence of treatment fidelity, quality of the outcome measures, and source of the report. Published appraisal tools intended specifically for single-subject research designs are a good option as well. Examples of these include the *"Evaluative Method"* from Reichow, Volkmar, and Cicchetti (2008), the revised *Risk of Bias in N-of-1 Trials* (RoBiNT) scale (Tate et al., 2013), or the *What Works Clearinghouse* standards (Kratochwill et al., 2010). Informally, you might consider the questions listed below when appraising single subject studies (Wendt & Miller, 2012).

1. Did the study include a sufficient number of baseline observations (at least five) and were the baseline observations relatively stable over time?
2. Did the researchers provide detailed, operational descriptions of the dependent measure(s), did they obtain repeated measurements in both baseline and experimental phases, and was the interobserver agreement for these measurements sufficiently strong?
3. Did the researchers provide detailed descriptions of the independent variable or experimental manipulation, and did they take steps to determine treatment fidelity and include that information in the research report?
4. Did the researchers demonstrate experimental control by showing a clearly identifiable change in behavior that occurred near the onset of treatment, and did they demonstrate this behavior change through repeated baseline and experimental phases?
5. For multiple baseline studies, did the researchers stagger the onset of treatment phases for new behaviors, participants, or settings?
6. Did the data analysis include graphs showing measurements in the baseline and experimental phases for visual inspection and also include statistical analysis to supplement visual inspection?

Table 10–1. Summary of Criteria for Critical Appraisal of a Research Report

1. Purpose or focus of the study		
2. Basic research design	Randomized group design	If yes, what type?
	Quasi-experimental group design	If yes, what type?
	Single participant design	If yes, what type?
	Case study	If yes, what type?
	Other	Specify
3. Other design features	Pretest	If yes, were participants similar?
	Long-term posttest	Briefly describe
	Blinding for outcome measurement	
4. Participants	Age	Briefly describe
	Diagnosis, if relevant	Briefly describe
	Gender	Briefly describe
	Cultural and linguistic background	Briefly describe
	Random assignment to groups	
	Random selection	If not randomly selected, briefly describe participant recruitment
	Number of participants in each group	
	Participant loss	If yes, did the authors include an explanation?
5. Treatment dependability/ fidelity	One or several clinicians	
	Qualifications of the clinician(s)	
	Procedures to monitor treatment fidelity	Briefly describe
6. Outcome measures	Tests with known reliability and validity	
	Informal or unique measures	If yes, did the authors include information about reliability and validity?
	Generalization measures	Briefly describe
	Measures of social importance (e.g., activity and participation)	Briefly describe

continues