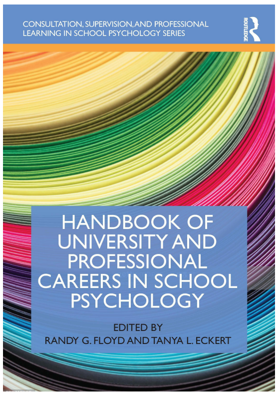

Handbook of University and Professional Careers in School Psychology

Randy G. Floyd, Tanya L. Eckert

Enhancing Skills in Research Methods and Statistics

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9780429330964-29
Nicholas F. Benson, David A. Klingbeil, Jacqueline M. Caemmerer
Published online on: 30 Dec 2020

20 Enhancing Skills in Research Methods and Statistics

Nicholas F. Benson, David A. Klingbeil, and Jacqueline M. Caemmerer

The goal of this chapter is to provide readers with information on how to establish and maintain a strong research foundation that incorporates high-quality research methods and sophisticated statistical techniques. While some readers may be mid-career or even later career faculty seeking to enhance their methodological and statistical expertise, the primary audience is advanced graduate students and early career scholars interested in acquiring requisite skills for the conduct of research and publication of findings. Thus, the logical first step for the primary audience is to identify essential skills that will be needed to thrive as they embark on their careers.

One approach for identifying methods and statistical techniques that are essential for those who study school psychology is to review the extant literature. However, while this approach provides information about research designs and analytic techniques that have been used, it does not directly inform decisions regarding designs and techniques that ought to be used. Thus, such results should not be interpreted as delimiting the skills researchers need but rather as highlighting those skills that have proven beneficial to myriad scholars who have published their work in school psychology journals. We reviewed all issues of the three highest impact school psychology journals published in the past 5 years, or more precisely between January of 2015 and October of 2019. These three journals include *Journal of School Psychology* (JSP, 2018 Impact Factor = 3.08), *School Psychology* (SP, 2018 Impact Factor = 2.08), and *School Psychology Review* (SPR, 2018 Impact Factor = 1.35). Of the 555 articles, most (472) were group-level quantitative studies. Comparatively, 31 were single-case experimental design studies (SCEDs); 22 were commentaries, including articles such as introductions to special issues and tutorials; 12 were systematic reviews; 10 were qualitative studies; and less than a handful were mixed method studies. The percent of publications during this time frame by research design is presented in Table 20.1.

Given the relatively high rate of studies with a group-level quantitative design, much of this chapter will focus on skills relevant to group-level quantitative research. However, there is undoubtedly a need for studies with SCEDs to help identify effective interventions as well as for qualitative and mixed method studies that describe the implementation of interventions or programs or help researchers generate research hypotheses. Moreover, many scholars who identify as school psychologists publish findings from SCEDs in related journals such as *Remedial and Special Education* and the *Journal of Behavioral Education*. Thus, we cover skills needed for SCEDs in depth, with the expectation that such studies are needed and that the number of such studies published in major school psychology journals will increase over time.

Group-Level Quantitative Research

The American Psychological Associations' Publications and Communications Board Task Force Report (Appelbaum et al., 2018) features journal article reporting standards (JARS) that specify information recommended for inclusion in quantitative research manuscripts. Authors,

Table 20.1 Designs Used in Recent Articles Published in Three Major School Psychology Journals

Type of Study	*% Articles Published*
Group-level quantitative	86
Single-case experimental design	6
Commentary	4
Systematic review	2
Qualitative design	2
Mixed method design	<1

as well as those scholars who contribute to the peer-review process for school psychology journals, routinely consult these standards to help ensure the transparency of published research articles. Among other information that researchers should report, the JARS highlight the need for descriptions of hypotheses and corresponding analytic strategies, sampling procedures, participants in the study, and the dataset. When clear descriptions are provided, it is easy to read articles and identify the research questions, design and method, and statistical analyses conducted. Likewise, most journal articles involving quantitative research now feature descriptive statistics as well as a matrix of correlation coefficients to aid the interpretation of results and facilitate secondary analyses.

Data Modeling and Inferential Statistical Techniques

Reschly and Ysseldyke (2002) contrasted the correlational and experimental traditions in psychology conceptualized by Cronbach (1957) and suggested that the former is the past and the latter the future. Techniques based on the general linear model, or the generalized linear model when analyzing variables with non-normal distributions, are routinely used for data modeling within many academic disciplines. About 51% of the group-level quantitative studies published in the past 5 years have involved regression analyses, including studies with observed variables only, latent variables (i.e., variables that cannot be directly observed but are inferred from observed indicators) only, or a combination of both. Many such studies have used structural equation modeling because it allows researchers to (a) test theory-based hypotheses, (b) simultaneously exam the direct and indirect effects of numerous predictors on numerous dependent variables, (c) account for measurement error, and (d) test measurement reliability. Although some experimental studies utilize regression models, this percentage undoubtedly indicates that the demise of the correlational tradition has been greatly exaggerated. Indeed, correlational data and regression techniques are routinely used for data modeling within many academic disciplines.

Analysis of variance (ANOVA), including extensions of this approach that blend ANOVA and regression to allow for inclusion of multiple dependent variables and/or covariates, were utilized in about 18% of the group-level quantitative studies published in the past 5 years. When multiple dependent variables are included in the model the analysis is referred to as MANOVA. When covariates (i.e., characteristics other than the treatment, or other independent variable of interest, that you want to control for) are included in the model the analysis is referred to as ANCOVA when there is a single dependent variable and MANCOVA when the model contains more than one dependent variable. Like regression, these modeling approaches are extensions of the general linear model, or the generalized linear model when working with non-normal data. While the ANOVA family of techniques may be used in studies that are not experiments, the frequency with which these techniques are used does highlight the importance of experimental studies in school psychology.

About 21% of the group-level quantitative studies that we reviewed used analyses to account for nesting effects in multilevel data structures. Multilevel modeling involves the use of regression equations to model parameters at more than one level. Nesting is a common issue in school psychology research, and the importance of addressing nested data is highlighted by the high citation rate of a practical guide to multilevel modeling published in JSP (Peugh, 2010). Based on our early 2020 extraction from the Scopus database of the most cited articles ever published in JSP, SP, or SPR, this article has been cited 522 times and ranks sixth among the most cited articles in these journals.

Researchers often need to address univariate or multivariate non-normality. About 13% of the group-level quantitative studies that we reviewed utilized distribution-free tests. Myriad others utilized either robust estimators to account for non-normality or utilized estimators that make no assumptions about the probability distributions of variables.

School psychologists frequently utilize tests and measures in their practice (Benson et al., 2019). Thus, it is not surprising that about 17% of the group-level quantitative articles we reviewed involved examining the structure/structural fidelity of tests and measures. Approximately 6% involved testing measurement invariance, a step that must be taken to ensure that the numerical values of test scores are on the same measurement scale across subgroups (e.g., age- or race-based groups). It is necessary to establish that test scores are not scaled differently or have different meanings based on group membership, whether scores are used for research purposes or for applied purposes.

About 11% of the group-level quantitative articles we reviewed involved testing mediators or moderators. Moderation and mediation can explain the “for whom” and the “how” of the relation between two variables (Fairchild & McQuillin, 2010). About 7% had longitudinal designs, with the typical aim of measuring and understanding growth or change. About 6% reported results from studies of diagnostic or classification accuracy. These studies examined the extent to which a test score, or set of test scores, has discriminative validity. That is, evidence supports the use of the score/s for differentiating diagnostic groups and indicates the scores can be used to classify individuals with adequate accuracy (Haynes, Smith, & Hunsley, 2019).

Finally, 6% of the group-level quantitative articles we reviewed were meta-analyses. While school psychology journals sometimes publish other types of systematic reviews (approximately 2% of articles based on our review of articles published in the past 5 years), meta-analysis is the most transparent approach because the “suppositions underlying inferences are made explicit, and therefore are open to public scrutiny and test” (Cook et al., 1992, p. 12). We proffer that most scholars interested in school psychology would benefit from obtaining the skills needed to complete meta-analytic studies. Data collection can be prohibitively expensive, in terms of both time and financial resources. Also, recent history has shown that global pandemics such as COVID-19 can disrupt collection of data from human subjects. Meta-analysis is often a sound alternative for addressing research questions when data collection is unfeasible. Further, meta-analysis helps to address limitations associated with individual studies and can be particularly useful when studying phenomena with low base rates.

Acquiring Quantitative Knowledge and Skills

Acquiring expertise regarding all available research methods and statistical techniques is seemingly an insurmountable task. Thus, focusing on that which has proven most useful is a reasonable strategy. One constant is the fact that researchers often work with datasets that are far from optimal. Thus, we begin by describing strategies for acquiring the skills needed to (a) screen and describe datasets, (b) evaluate and handle missing data, and (c) handle non-normal data. We then describe general knowledge and skills for data modeling and statistical inference. Next, we discuss specific issues germane to quantitative research in school psychology. In addition to the readings cited in the text, Table 20.2 includes a list of helpful resources related to these topics.

Screening and describing datasets. As Jacqueline's doctoral advisor, Timothy Z. Keith (University of Texas, Austin), reminded students before they began statistical analyses, "Always, always check your data!" This is a good reminder for all researchers, no matter the stage, as issues related to screening and describing data are frequently mentioned in reviewers' comments for journal article submissions. Run descriptive statistics, verify that values are within expected limits by examining minimum and maximum values, and check the amount of missingness for each variable. Verify the accuracy of the entered data against the measures used. For example, ensure items were reverse scored on a rating scale (if applicable). Ensure missing data are appropriately identified in the dataset. Some statistical software packages assign blank cells to missing data, while others require you to assign a value such as -9, -999, or any non-plausible value in your dataset. Test common statistical test assumptions (e.g., test normality by examining skewness and kurtosis) to determine whether any assumptions are violated, and which statistical test is most appropriate for your data. Visualizing the data may also be useful, such as running histograms and box plots. While examining your data, you may also need to create variables, such as composite scores or dummy variables for categorical data. The best way to develop these skills is to familiarize yourself with statistical software and practice, practice, practice! Table 20.2 includes a list of resources for data screening.

Carefully screening and setting up your data will save you precious research time. You may be so excited to begin statistical analyses and receive the results that you skip this important step. The risk is that you miss an anomaly in your data, and your resulting findings are not valid or your model may not run properly, leading to wasted time troubleshooting the error.

To best describe your data, familiarize yourself with the literature within your specialty and determine what information is often reported. Report major demographic characteristics of your sample (e.g., gender, race and ethnicity, and socioeconomic status of the sample and the school students attended) and characteristics of the sample relevant to your study such as average achievement levels.

Evaluating and handling of missing data. The JARS include a call for researchers to specify the extent of missing data in their study and how any missingness was handled. The importance of addressing missing data in school psychology research is highlighted by the high citation rate of an introduction to missing data published in the JSP (Baraldi & Enders, 2010). In early 2020 we extracted the most cited articles ever published in JSP, SP, or SPR from the Scopus database, and this introduction ranked fifth with 546 citations. Missing data are common in research, but missingness may affect statistical power due to smaller sample sizes and may introduce bias related to the cause of the missingness. Sound analysis requires that researchers evaluate mechanisms that underlie missingness to ensure that appropriate techniques are selected and used to handle missingness and minimize bias in results obtained from statistical analyses.

To conceptualize how missing data relate to the variables under study, Ruben and colleagues' (1976) delineated three missing data mechanisms: missing completely at random, missing at random, and missing not at random (Enders, 2010). Missing completely at random is optimal because data are considered unbiased; there is no association between the cause of the missingness and values on the variables under study. Little's missing completely at random test is available in many software packages to determine whether all the variables in your data meet this condition. With missing at random data, missingness is related to some variables but remains unrelated to values on the variables under study so long as auxiliary variables are controlled. Auxiliary variables are related to the missingness and assist with the estimation process but are not included in the model. Currently, there is no test for missing at random data. Finally, data that are missing not at random are related to values on the variables under study, even after controlling for other variables, and the results may be biased (Enders, 2010).

Researchers have several options when handling missing data. Listwise and pairwise deletion removes cases with missing data, which reduces your sample size and statistical power. Listwise and pairwise deletion will not bias results for missing completely at random data but may bias results for missing at random and missing not at random data (Schafer & Graham, 2002). Full-information maximum likelihood estimation and multiple imputation allow researchers to analyze data with incomplete data, which helps with power. Full-information maximum likelihood estimation is an iterative process that predicts the incomplete variables from the observed variables. When applied to missing completely at random and missing at random data, full-information maximum likelihood produces unbiased results (McArdle, 1994; Rubin, 1987). Unfortunately, full-information maximum likelihood estimation is unavailable in many software packages but is available in structural equation and multilevel modeling software. However, multiple imputation is more readily available in software packages. Multiple imputation replaces missing values with a set of estimated values based on several plausible alternative versions of complete datasets (Schafer & Graham, 2002). Drawbacks of multiple imputation include the use of inappropriate imputation models, and the uncertain imputed values are treated the same as the observed values in the analysis (Schafer & Graham, 2002). For a list of readings on missing data, see Table 20.2.

Handling non-normal data. While statistical tests can be invaluable tools, errors commonly occur due to violations of inherent assumptions. Parametric tests assume that data follow a normal distribution. There are numerous parametric tests (e.g., bivariate correlation, linear regression, *t*-tests, and ANOVA). Moreover, statistical techniques for modeling of sample variance and covariance matrices (e.g., factor analysis, structural equation modeling) most commonly

Table 20.2 List of Resources for Further Reading

1	University of California Los Angeles' Institute for Digital Research & Education (IDRE) website: https://stats.idre.ucla.edu. Topics include screening data, syntax for a variety of statistical software packages, evaluating and handling missing data, decision tree for selecting appropriate statistical test based on your data, multilevel modeling, RANOVA, mediated moderation, structural equation modeling, and many more. Search the seminars section for further reading.
2	YouTube videos and software publisher and universities websites: step-by-step directions for data screening.
3	Craig Enders' 2010 book, *Applied Missing Data Analysis*. Topics include missing data, handling missing data, maximum likelihood estimation, multiple imputation, Bayesian estimation, and much more.
4	Tim Keith's 2019 book, *Multiple Regression and Beyond*. Topics include confirmatory factor analysis, structural equation modeling, measurement invariance, latent growth analyses, mediation and moderation, regression, and much more.
5	John Loehlin's and Alex Beaujean's 2017 book, *Latent Variable Models*. Topics include exploratory and confirmatory factor analysis, structural equation modeling, path analysis, and much more.
6	Barbara Tabachnick's and Linda Fidell's 2014 book, *Using Multivariate Statistics*. Topics include factor analysis, ANOVAs, regression, screening data, and much more.
7	Kristopher Preacher's and Geoffrey Leonardelli's website, http://quantpsy.org/sobel/sobel.htm. An interactive calculation tool for mediation, helpful background information, and SAS and SPSS syntax are available.
8	Youngstrom's 2014 article is a useful primer for receiver operating characteristic analysis.
9	David Kenny's website, http://davidakenny.net/. Topics include structural equation modeling, mediation, moderation, and many more. Webinars, Power Points, and articles are available.
10	Linda and Bengt Muthén's Mplus (Muthén & Muthén, 2020) website, http://statmodel.com/. Topics include Mplus syntax, structural equation modeling, latent growth modeling, measurement invariance, mediation, Bayesian structural equation modeling, and many more. Handouts, webinars, and articles are available.

involve maximum likelihood estimation (Lei & Wu, 2012). When data are non-normal, maximum likelihood estimation tends to underestimate standard errors and inflate estimates of model fit. Thus, it is imperative to test univariate and multivariate normality to ensure that the data follow a normal distribution. Note that in many cases it is unnecessary for all variables to be normally distributed. For example, a grouping variable in ANOVA (e.g., a variable designating treatment or control group) obviously is not expected to be normally distributed.

Numerous nonparametric, also known as distribution-free, tests have been developed and can be used when it would be inappropriate to use parametric tests. Examples include Spearman's rank correlation coefficient, Kruskal-Wallis test, and statistical bootstrapping techniques. Also, variants of the maximum likelihood estimator have been developed for use with non-normal data (e.g., robust maximum likelihood estimation and full information maximum likelihood). Bayesian estimation is another option when working with non-normal data. Bayesian estimation does not make any restrictions about the form of data. Rather, the estimation process involves computing posterior distributions for parameters through the combination of prior distributions and empirical data (Lei & Wu, 2012).

Data modeling and statistical inference. Researchers are faced with the decision of which statistical test is most appropriate to answer their research questions. The scaling of the independent (also known as predictor) and dependent (also known as outcome) variables under study guide this decision. Generally, the ANOVA family of tests are used when the researcher is interested in group differences. The independent variables are categorical and the dependent variable is continuous. MANOVA tests two or more dependent variables (multivariate) that are correlated with each other, such as science and reading achievement. ANCOVA and MANCOVA control for covariate(s) that influence the dependent variable. An example of a covariate is previous achievement when studying later achievement. If relations between variables and prediction are the focus, regression may be a good choice. Simple or multiple regression (more than one independent variable) allow for dichotomous or continuous independents, and the dependent variable is continuous (Tabachnick & Fidell, 2019).

Relatedly, logistic regression and multilevel modeling can address regression-like research questions. Logistic regression is appropriate for dichotomous dependent variables. Multilevel modeling—also known as hierarchical linear modeling—accounts for the influence of clustering (e.g., students within schools) and estimates coefficients at the individual and cluster or group level (Keith, 2019). Path analysis is also an extension of regression and can answer questions about indirect and direct effects of independent variables. The introduction of latent, or unobserved, variables directs the researcher to structural equation modeling, particularly when there are multiple indicators of a construct and an adequate sample size (Keith, 2019).

Training and experience are other factors that guide decisions about statistical test selection. Most often researchers obtain the background knowledge for data modeling and statistical inference through formalized coursework, which begins with your first statistics and methods course in graduate school. Many training programs require a few statistics and methodological courses, but to better develop these skills, researchers should adventure beyond the required and complete additional optional coursework. Become familiar with the typical techniques used in your specialty and seek out related coursework; faculty can stress the importance of this to their mentees. In addition to coursework, seize opportunities to apply the statistical techniques in graduate school under the supervision of a faculty member or a more experienced graduate student. You may even request to reanalyze data they previously analyzed to practice skills and verify your accuracy in using new skills.

For those in postdoctoral and faculty positions, professional conferences sometimes host introductory statistical trainings and many university-based quantitative methods or statistics programs offer brief (1–2 weeklong) summer statistics courses. Similarly, the American Psychological Association, the Institute of Education Sciences, and other professional organizations offer summer training institutes for early, middle, and later career researchers. The focus of such trainings may be particular statistical techniques (e.g., structural equation modeling and multilevel modeling), research methodology (e.g., single-case experimental design), or software skill development (e.g., R). Some trainings require applications to participate and may offer financial assistance. Aside from formal training opportunities, consult and collaborate with others with more statistical and methodological experience to strengthen your own skills.

As with any other skill, knowledge is only part of the equation. After attending a statistical or methodological training, practice applying the technique to data while the information is still fresh in your mind in order to solidify the knowledge. Enhancing statistical and methodological skills is a continuous process. Even after you develop mastery in a technique, new developments will be published so you will need to stay up-to-date and evolve. Being an active consumer of researcher, as well as producer, will expose you to new developments.

Examining structural fidelity. It is imperative that appropriate criteria be used to evaluate tests and other assessment techniques to determine if there is adequate evidence to support interpretations and uses of test scores. There are obvious practical and legal–ethical problems with using poorly supported test scores for applied purposes such as diagnosis, prediction, or treatment planning. Likewise, good science will not result if inadequate and invalid measures are used in school psychology research. Meehl (1978) lamented the use of "step-wise low validation," whereas strong correlations (e.g., ≥.6) are used as evidence of validating a substantive theory; he described this as about as close to a "scientific nothing" as you can get (p. 823). Unfortunately, these types of weak validation studies remain far too common, and there is a strong need for researchers with measurement literacy to help ensure that the quality of research in this area will continue to improve.

The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) were developed to guide the process of developing, evaluating, and utilizing tests. The *Standards* summarize five categories of validity evidence: (a) evidence based on test content, (b) evidence based on response processes, (c) evidence based on internal structure, (d) evidence based on relations to other variables, and (e) evidence for validity and consequences of testing. Relative to other categories, articles that report evidence based on internal structure have been published far more frequently in the school psychology literature.

Internal structure refers to the number of latent variables (or factors) measured by tests and their pattern of relations with indicators (such as test items or subtests). Structural fidelity, a necessary condition for internal validity, exists when the internal structure corresponds adequately to the substantive domain of interest (Loevinger, 1957). Two common techniques for examining internal structure are exploratory and confirmatory factor analysis. Exploratory factor analysis is often used with newer, less understood measures as no factor structure is imposed. Researchers must make a number of decisions when using exploratory factor analysis, such as the method and criteria to decide how many factors to extract (e.g., principal component analysis, principal axis factor analysis, and maximum likelihood), rotation methods, and cut-off values for the magnitude of factor loadings (i.e., the strength of the relation of the item with the factor; Tabachnick & Fidell, 2019). Universal recommendations for these decisions do not exist, therefore remaining abreast of exploratory factor analysis developments within the methodological literature is key. Confirmatory factor analysis imposes a factor structure on the data, and thus are often used with established tests. Model fit indices suggest whether the imposed structure is a good fit or not to the data. Confirmatory factor analysis falls under the structural equation modeling umbrella (Keith, 2019). For a list of related readings, see Table 20.2.

Invariance testing. The *Standards* (AERA et al., 2014) also emphasize the importance of fairness in testing. While fairness is an abstract, social-political issue that is difficult to objectively evaluate, procedures have been developed that allow researchers to evaluate test bias (i.e., differential validity of test scores based on group membership). Measurement invariance testing allows researchers to examine if a test measures constructs similarly across gender, race and ethnicity, different ages, or any other grouping variable. This process involves adding increasingly stringent constraints across the groups. If the constraints are supported, they are retained in the following steps. The steps are, in order from least to most stringent: (1) configural, (2) metric (also known as weak invariance), and (3) and scalar (strong or intercept invariance; Meredith, 1993). Tests of configural invariance determine the extent to which the same factor structure fits the data across groups. In other words, if the items, subscales, or subtests (i.e., the observed variables that are directly measured) are associated with the same underlying latent variables/factors. Tests of metric invariance involve constraining factor loadings for each observed variable to be equal across groups. If the additions of these constraints do not degrade model fit, then it can be concluded that the scaling of latent variables is equal across groups (i.e., each unit change in the latent variable results in the same change in the observed variables used as indicators of the latent variable). Scalar invariance constrains factor loadings and intercepts equal across groups and indicates that the starting points of observed variables do not differ systematically across groups (Keith, 2019). A fourth step, strict or residual invariance, may be tested, but not all methodologists agree this is necessary (Vandenberg & Lance, 2000; Widaman & Reise, 1997). Strict invariance requires factor loadings, intercepts, and residual variances to be constrained across groups and suggests differences in the means and variances of the measured variables are fully explained by differences in the latent variable means and variances (Keith, 2019). Fit indices are examined to determine if invariance is supported.

Invariance testing itself may be the focus of a research study, or invariance testing may be a necessary precursor before a broader analysis can be conducted. Measurement invariance is required before latent means and variances can be compared across groups in structural equation modeling. Metric invariance is needed before comparisons in paths (the influence of one latent variable on another) can be made across groups (Keith, 2019). Additionally, predictive validity can be examined to determine if a latent variable predicts one or more outcome variables similarly across groups.

Examining mediation and moderation. As previously noted, moderation and mediation can explain the "for whom" and the "how" of the relation between two variables; both are sometimes referred to as third variable effects models (Fairchild & McQuillin, 2010). Such studies tend to advance both theory and clinical practice by addressing (a) important theoretical and practical questions regarding "how" (i.e., mediation) and "when"' or "for whom" (i.e., moderation) effects occur and (b) how these processes may be integrated (i.e., mediated moderation and moderated mediation) in real-world phenomena (Karazsia, Berlin, Armstrong, Janicke, & Darling, 2014). Publication of numerous studies addressing these processes can be viewed as a hallmark of a maturing profession (e.g., Judd, McClelland, & Culhane, 1995).

Moderation and interaction are synonymous. In moderation the magnitude of the effect of an independent variable on the dependent variable depends on the level of the moderator ("for whom"). For example, the influence of working memory on sentence composition is stronger for younger children than older youth (Caemmerer, Maddocks, Keith, & Reynolds, 2018). Mediation and indirect effects are synonymous. Mediator variables either partially or completely explain the influence of an independent variable on a dependent variable (the "how or why"). In mediation, (1) the independent variable statistically significantly influences the dependent variable (although some argue this may not be necessary), (2) the independent variable significantly influences the mediator, (3) the mediator significantly influences the mediator variable, and (4) the influence of the independent variable on the dependent variable is reduced when the mediator is included (see Table 20.2 for an online resource).

Mediation and moderation can be integrated, referred to as moderated mediation and mediated moderation, because the two approaches overlap. In moderated mediation, also referred to as a conditional indirect effect, the influence of a mediator varies depending on the level of a moderator. An example is the influence of child body-mass index on child body dissatisfaction, which is mediated by mother's encouragement to diet; this mediator is moderated by mother's own body-mass index (Karazsia et al., 2014). In mediated moderation, the strength of the moderator, or interaction effect, can be explained by a mediator.

Examining classification/diagnostic accuracy. Practical matters such as screening and diagnosis involve classifying individuals into categories. These matters often rely upon the establishment of cut-off scores to indicate at-risk status or clinical concern. Continuous data tend to be unreliable when dichotomized as falling either at or above a cut-off score or below a cut-off score (e.g., MacCallum, Zhang, Preacher, & Rucker, 2002). Given limitations with the use of cut-off scores, the accuracy of classification decisions must be evaluated closely. Diagnostic accuracy statistics can be used to index how well tests discriminate between individuals with and without a target condition. Several online calculators are available to calculate these statistics. Also, receiver operating characteristic analysis can be used to evaluate the accuracy of decisions. Receiver operating characteristic analysis has the important practical advantages of quantifying the trade-off between sensitivity and specificity (i.e., area under the curve) and identifying optimal cut scores for classification systems. Recommended reading is listed in Table 20.2.

Modeling change using longitudinal designs. Often researchers are interested in studying change or growth in a behavior or characteristic over time. Longitudinal research includes repeated measurements at two or more time points. Like studies of mediators and moderators, longitudinal research has the potential to advance theory as well as clinical practice. Rationales for conducting longitudinal research include but are not limited to the following: (a) studying intraindividual change in attributes over time, (b) studying interindividual differences in intraindividual change, (c) studying interrelationships among multiple attributes that may occur simultaneously and/or sequentially, (d) analyzing causes (determinants) of intraindividual change, and (e) analyzing causes (determinants) of interindividual differences in intraindividual change (Baltes & Nesselroade, 1979).

There are several statistical options for analyzing change over time. Repeated measures (i.e., within-subjects) ANOVA compares means on categorical independent variables. For example, you may implement a depression treatment and assess children's symptoms severity at 2, 4, and 6 weeks. Multilevel modeling can also be used to analyze repeated measures. In multilevel modeling repeated measures are nested within individuals; time is nested within persons. There are several different approaches to analyze growth within structural equation modeling (Keith, 2019). Longitudinal panel models examine the interrelationships, or reciprocal relations, between change in behaviors over time. For example, children's social skills influence their subsequent academic achievement and vice versa from kindergarten to fifth grade (Caemmerer & Keith, 2015).

Latent growth models test initial levels (i.e., intercept) of a behavior and growth (i.e., slope) over time. Additionally, latent growth models assess average growth over time (intraindividual change) as well as the amount of variation in growth across children over time (interindividual variation). Latent growth models can include covariates that influence individuals' growth such as socioeconomic status. Latent growth models can be extended to involve multigroup analyses, multiple indicators of constructs (curve of factors), different trajectories for subgroups of individuals (latent class growth analysis), and the interplay of growth between two or more attributes simultaneously (bivariate or multivariate latent growth models; Little, 2013). When multiple observed variables are used as indicators of an attribute, longitudinal invariance should be examined to determine whether the same construct is measured at each time point (Widaman, Ferrer, & Conger, 2010). Table 20.2 includes a list of related resources. Nicholas attended the American Psychological Association's Advanced Training Institute on Structural Equation Modeling in Longitudinal Research and highly recommends it to anyone interested in longitudinal designs.

Systematic Reviews and Meta-Analysis

Although systematic reviews and meta-analyses account for only about 7% of articles published in school psychology journals, there has been rapid growth in the number of published meta-analyses in related fields such as education, psychology, and medicine (Shadish & Lecy, 2015). The popularity of meta-analysis seems likely to continue given the current focus on identifying evidence-based practices and answering questions about what practices work for whom and under what conditions (e.g., Fuchs & Fuchs, 2019; Ledford et al., 2018). We certainly expect that readers are used to seeing meta-analyses in the journals they typically read including those in school psychology. Despite the popularity of meta-analysis, published research syntheses certainly vary in statistical and methodological rigor (Ahn, Ames, & Myers, 2012; Tipton, Pustejovsky, & Ahmadi, 2018).

Basic considerations. Meta-analysis generally refers to methods of combining effect size estimates from multiple primary studies to (a) estimate an average treatment effect, (b) evaluate the consistency of treatment effects, (c) evaluate the effect of explanatory variables that differ across studies, and (d) test the significance of variation in treatment effects across levels of the explanatory variables (Hedges & Olkin, 1985). There have been substantial, untold advances in meta-analytic methods since the seminal work of Glass (1976) who pioneered a method for combining results of studies with different dependent variables. Still, the basic steps of conducting a meta-analysis are relatively straightforward. The steps include formulating a problem, searching the literature and collecting research evidence, analyzing the evidence from the primary studies, interpreting the results, and disseminating the findings (Cooper & Hedges, 2009). Meta-analytic methods can be applied to outcomes that are naturally dichotomous (e.g., relative risk or odds ratios), mean differences (e.g., raw or standardized mean differences), or correlation coefficients that analyze the association between two variables (Borenstein, Hedges, Higgins, & Rothstein, 2009). A full description of these steps is beyond the scope of this chapter. Instead, our goal is to provide interested readers with helpful starting points for learning how to conduct high-quality meta-analyses.

Helpful guides for unfamiliar researchers. There are several textbooks, introductory articles, and special issues which provide clear introductions and step-by-step directions to conducting a meta-analysis (see Table 20.3). David became interested in the topic when discussing a potential topic for meta-analysis with colleagues. Although David had read published meta-analyses, the process for completing one became much clearer after reading the Borenstein et al. (2009) textbook, which he selected primarily because his university library had the e-book available for free. He found the text to be easily consumable and to provide a foundation to (a) better understand published meta-analyses, (b) better understand more complex texts that are heavier on statistics, and (c) identify questions that he needed to answer when setting out to conduct his own meta-analysis. Methodological guidance can also in top-tier journals such as *Psychological Methods, Statistics in Medicine, Research Synthesis Methods,* and *Review of Educational Research.*

Another potentially helpful source for hopeful meta-analysts are published articles that describe current best practices and standards for the conduct and reporting of meta-analytic findings. Pigott and Polanin (2020) provided a comprehensive review of the elements required of a high-quality meta-analysis. Tipton et al. (2019) provided a similar review of the elements required for conducting high quality meta-regression analyses. In addition to the reviewing these types of resources, it may also be helpful to review published standards for meta-analyses (Appelbaum et al., 2018; Higgins et al., 2019). By identifying what is required for conducting and reporting a high-quality meta-analysis, reading these standards may help prospective meta-analysts identify procedures they need guidance on. For example, the APA JARS (Appelbaum et al., 2018, p. 22) requires authors to describe (a) "methods used to assess risk to internal validity in individual study results" and (b) "the model used to estimate the heterogeneity of the effects sizes (e.g., a fixed-effect, random effect robust variance estimation)." Someone interested in learning how to conduct a meta-analysis would, at the minimum, know they would need to learn about the difference between fixed and random effect models as well as what robust variance estimation is. Becoming familiar with best practices in meta-analysis may also be helpful for consuming published meta-analyses in your research area of interest.

Software for conducting meta-analysis. A hallmark of meta-analysis is the quantitative synthesis of effect sizes extracted from primary studies (Hedges & Olkin, 1985). Conducting a meta-analysis will require the use of statistical software. Some specialized software for conducting meta-analysis is free (e.g., RevMan), while other software must be purchased (e.g., Comprehensive Meta-Analysis; Borenstein et al., 2013). Meta-analyses can also be conducted in popular statistical analysis programs such as SAS (SAS Institute Inc., 2018), Stata (StataCorp., 2019), R (R Core Team, 2019; Polanin, Hennessy, & Tanner-Smith, 2017), and Microsoft Excel (Suurmond, van Rhee, & Hak, 2017). Readers should note that some of these programs vary in their offerings.

Some programs allow users to calculate an effect size across a number of different formats. Primary studies often vary in the information that is reported. The Comprehensive Meta-Analysis program, as an example, allows users to estimate an effect size (e.g., Hedges' g) regardless of whether the primary study reported (a) group means, standard deviations, and sample sizes; (b) the overall sample size and the F statistic from a test of between-group changes; or (c) a number of other possible formats. Other programs and packages only allow users to conduct the actual meta-analytic synthesis using a file of effect sizes that have already been estimated. Readers interested in using a specific program are encouraged to search for guides on how to conduct a meta-analysis using that specific software as a number of these guides exist (see Table 20.3).

When conducting meta-analysis of between-group designs, David has used the Comprehensive Meta-Analysis software to extract effect sizes due to the extensive number of formats you can enter data to obtain the effect size estimates (e.g., Hedges' g and product-moment correlations). He then saved those data into a.csv file and conducted the actual meta-analytic analyses in R, primarily using the metafor and robumeta packages as they provide more flexibility in conducting the analysis. He supposes that readers who are more skilled in the use of R would find the step of using Comprehensive Meta-Analysis to calculate the effect sizes unnecessary.

Single-Case Experimental Designs

Single-case experimental designs (SCEDs) provide a rigorous technology to evaluate the functional relation between an independent variable (e.g., an academic or behavioral intervention) and a dependent variable (Kennedy, 2005). Although the prevalence of SCEDs in our review was low (6%), prior reviews of school psychology journals suggest that SCEDs represent a sizeable proportion of articles describing intervention research (e.g., Burns et al., 2012). The nature of SCEDs (e.g., small sample sizes) may increase their feasibility for researchers without substantial resources (e.g., grant funding) and perhaps their acceptability to schools in comparison to randomized controlled trials. In addition, SCEDs appear to fit well within the context of multi-tiered systems of support wherein educators are tasked with providing supplemental services to students in need of additional support (Skinner, McCleary, Skolits, Poncy, & Cates, 2013). Still, for researchers who are unfamiliar with SCEDs, the feasibility and potential acceptability of these designs should not be mistaken for ease. Fortunately, there are several guides and strategies for acquiring the knowledge for implementation of SCEDs. In the following sections, we describe basic considerations for the use of SCEDs followed by several resources we believe would be helpful to researchers who are unfamiliar with SCEDs.

Table 20.3 List of Resources Regarding Systematic Reviews and Meta-Analysis

1	Michael Borenstein's and colleagues' (2009) book, *Introduction to Meta-Analysis*. Topics include methods of estimating effect sizes, fixed and random effects models, quantifying heterogeneity, meta-regression, handling complex data structures, why and when to conduct a meta-analysis, and reporting.
2	Harris Cooper's book (2015), *Research Synthesis and Meta-Analysis*. Topics include formulating a research question, searching the literature, retrieving information from primary studies, evaluating the quality of studies, analyzing and synthesizing study outcomes, and interpreting and presenting the results.
3	Frank Schmidt's and John Hunter's (2015) book, *Methods of Meta-Analysis*. Topics include locating, evaluating, and coding studies, integrating research findings across studies, study artifacts and their impact on study outcomes, meta-analysis of correlations, experimental effects and other comparisons, publication bias, and meta-analysis methods and related software. This text is commonly cited for their discussion of correcting effect sizes (e.g., correlations) based on study artifacts.
4	William Shadish's (2014) paper, "Analysis and Meta-Analysis of Single-Case Designs: An Introduction," published in the *Journal of School Psychology*. This introduction to a special issue on the topic of meta-analysis of single-case designs provides an overview of the considerations that must be made for synthesizing results from single-case experimental designs.
5	Terri Pigot's and Joshua Polanin's (2020) paper, "Methodological Guidance Paper: High-Quality Meta-Analysis in a Systematic Review" published in *Review of Educational Research*. This article discusses the topics including research questions suited for meta-analysis, conducting quality search and coding procedures, publishing a review protocol, handling missing data, reporting the effect sizes and heterogeneity, handling dependent effect sizes, and exploring the impact of publication bias. The authors also provide guidance for interpretation and reporting of results. Notably, the authors highlight the use of meta-regression using robust variance estimation for handling dependent effect sizes and exploring multiple potential moderators.
6	Elizabeth Tipton's and colleagues' (2019) paper, "Current Practices in Meta-Regression in Psychology, Education, and Medicine" published in *Research Synthesis Methods*. This paper summarizes trends in rigorous journals publishing meta-analyses and discusses important gaps between methodological guidance and published meta-analytic syntheses.
7	Joshua Polanin's and colleagues' (2017) paper, "A Review of Meta-Analysis Packages in R," published in *the Journal of Educational and Behavioral Statistics*, reviewed the numerous meta-analytic packages in R and their specific features. They also provide a tutorial of two of the popular packages, metafor and robumeta.
8	Emily Tanner-Smith's and Elizabeth Tipton's (2014) paper, Robust Variance Estimation With Dependent Effect Sizes: Practical Considerations Including a Software Tutorial in Stata and SPSS," published in *Research Synthesis Methods*, provides a tutorial for conducting meta-regression using robust variance estimation in two popular statistical analysis packages.
9	Emily Tanner-Smith's and colleagues' (2016) paper, "Handling Complex Meta-Analytic Data Structures Using Robust Variance Estimates: A Tutorial in R," published in the *Journal of Developmental and Life-Course Criminology*, provides step-by-step directions to use the robumeta package in R. The article includes accompanying R code that readers could replicate to conduct their own meta-analyses.

Basic Considerations

The underlying logic of single-case research requires the repeated measurement of a dependent variable for each case and that each individual case provides its own control. A functional (i.e., causal) relation between an independent variable and a dependent variable is established when (a) a change in the dependent variable is observed when the independent variable is manipulated (e.g., intervention implementation or cessation) and (b) the change in the dependent variable is replicated at different times (Horner et al., 2005; Ledford, 2018). Current standards suggest that a change associated with manipulation of an independent variable (i.e., a basic effect) must be replicated at three different times before a functional relation can be inferred (Kratochwill et al., 2010). Therefore, researchers considering the use of SCEDs must ensure that the dependent variable of interest can be reliably and validly measured repeatedly over time.

Researchers must also determine the type of design that is best suited to answering their research question (Horner & Odom, 2014). SCEDs include a number of different designs including reversal designs, multiple-baseline and multiple probe designs, and alternating treatment designs. In fact, these types of designs can be combined to answer multiple research questions.

There is a long-standing tradition of graphing data in time-series fashion and visually analyzing the results in SCED research (Kratochwill et al., 2010). Briefly, researchers evaluate the level, trend, and variability of data within each phase. Differences across phases are evaluated based on how immediate the change is once the independent variable is manipulated (i.e., immediacy), the amount of overlap in the data between phases, and the consistency of the observed changes in the dependent variable across each phase change (Ledford, 2018). The statistical analysis of SCED results as a supplement to visual analysis is also increasing in frequency. Researchers have a number of options for estimating intervention effects in SCEDs quantitatively and there is no widespread consensus regarding the most appropriate option.

Helpful Guides for Unfamiliar Researchers

There are several books that provide comprehensive guidance regarding the planning and design, conduct, and analysis of SCEDs (see Table 20.4). There are also a number of peer-reviewed articles describing specific design types (e.g., Hartmann & Hall, 1976; Horner & Baer, 1978); considerations for the graphing and presentation of SCED data (Barton & Reichow, 2012; Radley, Dart, & Wright, 2018), the visual analysis of SCED data (e.g., Lane & Gast, 2014), and the application of SCEDs to specific types of interventions (e.g., Eckert, Ardoin, Daisey, & Scarola, 2000) or specific dependent variables (e.g., Klingbeil, Van Norman, & Nelson, 2017; Rizvi & Nock, 2008). David used programs such as Powerpoint or Excel to construct graphs until he had colleagues recommend GraphPad Prism. This program must be purchased; however, David

Table 20.4 List of Resources Regarding Single-Case Experimental Designs

1	Alan Kazdin's (2011) text, *Single-Case Research Designs*. Topics include measurement and assessment considerations, evaluating results, and displaying single-case data. There are chapters devoted to types of experimental designs (e.g., reversal, multiple-baseline, and combined designs), which describe the types of inferences allowed, situations in which the designs are appropriate, and general design considerations. This text has less considerations for the study of academic skill outcomes than the following two resources.
2	Craig Kennedy's (2005) text, *Single-Case Design for Educational Research*. Topics include hallmarks of single-case experimental designs (e.g., functional relations and direct and systematic replication), measuring behavior, and data-analysis. There are chapters devoted to different types of designs (e.g., reversal, multiple-baseline, and combined designs), which describe the types of inferences allowed, situations in which the designs are appropriate, and general design considerations.
3	Jennifer Ledford's & David Gast's (2019) text, *Single-Case Research Methodology: Applications in Special Education and Behavioral Sciences*. Topics include replication, writing tasks, selection of dependent and independent variables, visual representation of data and visual analysis, quality and rigor indices for single-case designs, and synthesis and meta-analysis of single-case research. There are chapters devoted to different types of designs (e.g., reversal, multiple-baseline and multiple-probe, combined, and comparative) that describe the types of inferences allowed, situations in which the designs are appropriate, and guidance regarding the implementation of the specific design type.
4	GraphPad Prism. This software provides several helpful options for graphing results from single-case experimental designs. Users can purchase an annual single license ($168 at the time this chapter was written), and there are discounts for student licenses ($108).

found it to be much more flexible and easier to use for graphing SCED data than Powerpoint or Excel.

Another set of resources that could be helpful to unfamiliar researchers are published design or reporting standards for SCEDs. The What Works Clearinghouse (WWC) pilot standards for SCEDs (Kratochwill et al., 2010) provide specific recommendations regarding several methodological aspects of SCEDs. The WWC standards provide recommendations regarding the number of phases, data points per phase, and collection of interobserver agreement data for reversal, multiple-baseline, and alternating treatment designs. Other standards such as the *Single-Case Reporting Guidelines in Behavioral Interventions* (Tate et al., 2016) and the Council for Exceptional Children's (2014) *Standards for Evidence-Based Practices in Special Education* provide detailed guidelines regarding the information researchers should report in manuscripts describing SCED research findings. Although reporting guidelines or standards generally provide less step-by-step guidance than the books have highlighted, reviewing them while planning an SCED may be helpful. For example, standard 3.2 in the Council for Exceptional Children (2014) standards indicates that a published study should describe the specific training or qualifications required to implement the intervention. If a researcher was planning a study where graduate students would be delivering the intervention, reviewing this standard could provide a helpful reminder that the details of the interventionist training should be carefully recorded.

A Call for Collaboration

It is impossible to capture all the nuances of conducting rigorous SCEDs in a section of a single chapter. Reading the textbooks, peer-reviewed articles, and published design standards may provide a solid foundation for researchers wishing to enhance their skills in conducting SCEDs. However, overlooking any of the critical aspects in conducting SCED research could result in a study that is fatally flawed and unpublishable. In order to (hopefully) avoid this frustrating result, we suggest that researchers who are conducting SCEDs for the first time consult with experts in this methodology while planning their study. Consulting with researchers with expertise in Applied Behavior Analysis and SCED methods, even if they are unfamiliar with the intervention or outcomes being studied, may help identify potential challenges or alternatives that are better suited to answer the research questions. These could be colleagues in special education or applied areas of psychology at the same institution or researchers at other institutions who have that expertise. Collaborating with other early career or mid-career faculty members who have a background in conducting SCEDs could result in a fruitful partnership for all parties involved.

Summary

The goal of this chapter is to provide readers with information on how to establish and maintain a strong research foundation that incorporates high-quality research methods and sophisticated statistical techniques. To this end, we presented strategies that can be used to acquire essential skills for quantitative research in school psychology. We first reviewed the extant literature and identified the research designs and analytic techniques most frequently used in studies published during the previous 5 years. We found that most of these studies followed the correlational tradition. In contrast to Reschly and Ysseldyke's (2002) prognostication, correlational research remains integral to the conduct of research in school psychology, likely because it is often impractical, unethical, or impossible to manipulate independent variables. Additionally, it is important to note that correlational studies are certainly appropriate when studying relationships among variables with no assumption of causality.

Although the correlational tradition is alive and well, it should be noted that the number of published intervention studies has increased dramatically since the turn of the 21st century

(Burns, Klingbeil, Ysseldyke, & Petersen, 2012). Moreover, frequency of use does not directly inform decisions regarding the designs and techniques that ought to be used. The usefulness of research is dependent on the research questions addressed as well as characteristics of the data analyzed. Optimally, research designs and statistical analyses should align with those research questions that can be tested to help advance the science and practice of school psychology.

Although school psychologists continue to spend the majority of their professional time administering tests and otherwise participating in activities germane to identifying students in need of special education (Benson et al., 2019), it appears that the frequency with which they engage in the broader range of services specified in the National Association of School Psychologists' Practice Model is increasing (McNamara, Walcott, & Hyson, 2019). Moreover, many school psychologists now report using problem-solving models and response-to intervention (RtI) data to help identify specific learning disabilities (Benson et al., 2020). Thus, those who engage in school psychology research would be wise to develop expertise pertaining to experimental designs and SCED in addition to developing quantitative skills from the correlational tradition.

We hope that our chapter provides a useful resource to those wishing to enhance their skills in research methods and statistics. As noted earlier, acquiring expertise regarding all available research methods and statistical techniques is seemingly an insurmountable task, so we focused primarily on skills that have proven useful for publishing in the school psychology. Our recommendations were designed to help readers develop skills consistent with best practices in research design and data analysis. Notably, methodological and statistical advances occur frequently, and thus obtaining expertise is akin to hitting a moving target. To remain productive, researchers need to stay abreast of current literature and be proactive in maintaining their knowledge and skills in research methods and statistics.

References

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Education Research, 82,* 436–476. https://doi.org/10.3102/0034654312458162

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73,* 3–25.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York, NY: Academic Press.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*, 5–37.

Barton, E. E., & Reichow, B. (2012). Guidelines for graphing data with Microsoft® Office 2007™, Office 2010™, and Office for Mac™ 2008 and 2011. *Journal of Early Intervention, 34*(3), 129–150.

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 national survey. *Journal of School Psychology, 72,* 29–48.

Benson, N. F., Maki, K. E., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Kranzler, J. K., & Fefer, S. A. (2020). A national survey of school psychologists' practices in identifying specific learning disabilities. *School Psychology, 35*, 146–157.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, UK: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P, & Rothstein, H. R. (2013). *Comprehensive Meta-Analysis*, Version 3. Englewood, NJ: Biostat.

Burns, M. K., Klingbeil, D. A., Ysseldyke, J. E., & Petersen, B. S. (2012). Trends in methodological rigor in intervention research published in school psychology journals. *Psychology in the Schools*, *49*, 843–851.

Caemmerer, J. M., & Keith, T. Z. (2015). Longitudinal reciprocal effects of social skills and achievement from kindergarten through eighth grade. *Journal of School Psychology, 53*, 265–281. doi: 10.1016/j.jsp.2015.05.001

Caemmerer, J. M., Maddocks, D. L. S., Keith, T. Z., & Reynolds, M. R. (2018). Effects of cognitive abilities on child and youth academic achievement: Evidence from the WISC-V and WIAT-III. *Intelligence, 68*, 6–20. doi: 10.1016/j.intell.2018.02.005

Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., . . . Mosteller, F. (1992). *Meta-analysis for explanation: A casebook*. New York, NY: Russell Sage Foundation.

Cooper, H. (2015). *Research synthesis and meta-analysis: A step-by-step approach*. Thousand Oaks, CA: Sage.

Cooper, H., & Hedges, L. V. (2009). Research synthesis as a scientific process. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 3–16). New York, NY: Russell Sage Foundation.

Council for Exceptional Children (2014). Standards for evidence-based practices in special education. Arlington, VA: Author. Retrieved from: https://www.cec.sped.org/~/media/Files/Standards/Evidence/based/Practices/and/Practice/EBP/FINAL.pdf

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671–684.

Eckert, T. L., Ardoin, S. P., Daisey, D. M., & Scarola, M. D. (2000). Empirically evaluating the effectiveness of reading interventions: The use of brief experimental analysis and single-case designs. *Psychology in the Schools, 37*(5), 463–473. https://doi.org/10.1002/1520-6807(200009)37:5%3C463::AID-PITS6%3E3.0.CO;2-X

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Fairchild, A. J., & McQuillin, S. D. (2010). Evaluating mediation and moderation effects in school psychology: A presentation of methods and review of current practice. *Journal of School Psychology, 48*, 53–84.

Fuchs, D., & Fuchs, L. S. (2019). On the importance of moderator analysis in intervention research: An introduction to the special issue. *Exceptional Children, 85*, 126–128. https://doi.org/10.1177/0014402918811924

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8. https://doi.org/10.2307/1174772

Hartmann, D. P., & Hall, R. V. (1976). The changing criterion design. *Journal of Applied Behavior Analysis, 9*(4), 527–532. https://doi.org/10.1901/jaba.1976.9-527

Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2019). *Scientific foundations of clinical assessment* (2nd ed.). New York, NY: Routledge.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press, Inc.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). https://doi.org/10.1002/9781119536604

Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: A variation of the multiple baseline. *Journal of Applied Behavior Analysis, 11*, 189–196. https://doi.org/10.1901/jaba.1978.11-189

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179. https://doi.org/10.1177/001440290507100203

Horner, R. H., & Odom, S. L. (2014). Constructing single-case research designs: Logic and options. In T. R. Kratochwill & J. R. Levin (Eds.). *Single-case intervention research: Methodological and statistical advances* (pp. 27–51). Washington, DC: American Psychological Association. https://doi.org/10.1037/14376-002

Judd, C. M., McClelland, G. H., & Culhane, S. E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology, 46*, 433–465.

Karazsia, B. T., Berlin, K. S., Armstrong, B., Janicke, D. M., & Darling, K. E. (2014). Integrating mediation and moderation to advance theory development and testing. *Journal of Pediatric Psychology*, *39*, 163–173.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.

Keith, T. Z. (2019). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (3rd ed.). New York, NY: Routledge.

Kennedy, C. H. (2005). *Single-case designs for educational research.* Boston, MA: Pearson/Allyn and Bacon.

Kenny, D. (2014, July 23). Welcome to David Kenny's website. Retrieved from http://davidakenny.net/

Klingbeil, D. A., Van Norman, E. R., & Nelson, P. M. (2017). Precision of curriculum-based measurement reading data: Considerations for multiple-baseline designs. *Journal of Behavioral Education, 26*, 433–451. https://doi.org/10.1007/s10864-017-9282-7

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from http://ies.ed.gov/ncee/pdf/wwc_scd.pdf

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24*(3–4), 445–463. https://doi.org/10.1080/09602011.2013.815636

Ledford, J. R. (2018). No randomization? No problem: Experimental control and random assignment in single-case research. *American Journal of Evaluation, 39*, 71–90.

Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology: Applications in special education and social sciences.* New York, NY: Routledge.

Ledford, J. R., King, S., Harbin, E. R., & Zimmerman, K. N. (2018). Antecedent social skills interventions for individuals with ASD: What works, for whom, and under what conditions? *Focus on Autism and Other Developmental Disabilities, 33*(1), 3–13. https://doi.org/10.1177/1088357616634024

Lei, P. W., & Wu, Q. (2012). Estimation in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 1640180). New York, NY: Guilford.

Little T. D. (2013). *Methodology in the social sciences. Longitudinal structural equation modeling.* New York: Guilford Press.

Loehlin, J. C., & Beaujean, A. A. (2017). *An introduction to factor, path, and structural equation analysis.* New York, NY: Routledge.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635–694.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research, 29*, 409–454.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

McNamara, K. M., Walcott, C. M., Hyson, D. (2019). Results from the NASP 2015 membership survey, part two: Professional practices in school psychology. *NASP Research Reports, 4*, 1–14.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543.

Muthén & Muthén (2020, March). *Mplus.* Retrieved from http://statmodel.com/

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*, 85–112.

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research, 90*, 24–46. https://doi.org/10.3102/0034654319877153

Polanin, J. R., Hennessy, E. A., & Tanner-Smith, E. E. (2017). A review of meta-analysis packages in R. *Journal of Educational and Behavioral Statistics, 42*(2), 206–242. https://doi.org/10.3102/1076998616674315

Preacher, K. J., & Leonardelli, G. J. (2010–2020). *Calculation for the Sobel test: An interactive calculation tool for mediation tests.* Retrieved from http://quantpsy.org/sobel/sobel.htm

R Core Team (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Radley, K. C., Dart, E. H., & Wright, S. J. (2018). The effect of data points per x- to y-axis ratio on visual analysts evaluation of single-case graphs. *School Psychology Quarterly, 33,* 314–322. https://doi.org/10.1037/spq0000243

Reschly, D. J., & Ysseldyke, J. E. (2002). Paradigm shift: The past is not the future. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (4th ed., pp. 3–20). Bethesda, MD: National Association of School Psychologists.

Rizvi, S. L., & Nock, M. K. (2008). Single-case experimental designs for the evaluation of treatments for self-injurious and suicidal behaviors. *Suicide and Life-Threatening Behavior, 38*, 498–510.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York, NY: John Wiley & Sons.

SAS Institute Inc. (2018). *SAS statistical software: Release 15.1.* Cary, NC: SAS Institute Inc.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177. doi: 10.1037/1082–989X.7.2.147.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis.* Thousand Oaks, CA: Sage Publications Inc.

Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology, 52*, 109–122. https://doi.org/10.1016/j.jsp.2013.11.009

Shadish, W. R., & Lecy, J. D. (2015). The meta-analytic big bang. *Research Synthesis Methods, 6*, 246–264. https://doi.org/10.1002/jrsm.1132

Skinner, C. H., McCleary, D. F., Skolits, G. L., Poncy, B. C., & Cates, G. L. (2013). Emerging opportunities for school psychologists to enhance our remediation procedure evidence base as we apply response to intervention. *Psychology in the Schools, 50*, 272–289. https://doi.org/10.1002/pits

StataCorp. (2019). *Stata statistical software: Release 16.* College Station, TX: StataCorp LLC.

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of meta-essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods, 8*, 537–553. https://doi.org/10.1002/jrsm.1260

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Boston, MA: Pearson/Allyn & Bacon.

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods, 5*(1), 13–30. https://doi.org/10.1002/jrsm.1091

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology, 2,* 85–112. https://doi.org/10.1007/s40865-016-0026-5

Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., Shadish, W., Horner, R., Kratochwill, T., Barlow, D. H., Kazdin, A., Sampson, M., Shamseer, L., & Vohra, S. (2016). The single-case reporting guideline in Behavioural interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology, 4*, 10–31. http://dx.doi.org/10.1037/arc0000027

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019). Current practices in meta-regression in psychology, education, and medicine. *Research Synthesis Methods, 10*, 180–194. https://doi.org/10.1002/jrsm.1339

University of California Los Angeles Institute for Digital Research & Educational Statistical Consulting (2020). Retrieved from https://stats.idre.ucla.edu/

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*, 10–18.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*, *39*(2), 204–221.