11 Identifying and Using Secondary Datasets to Answer Policy Questions Related to School-Based Counseling: A Step-by-Step Guide

Julia Bryan, Jungnam Kim, and Qi Shi

Identifying and Using Secondary Datasets to Answer Policy Questions Related to School-Based Counseling: A Step-by-Step Guide

Researchers have demonstrated the pivotal role that school counseling plays in addressing the academic, socio-emotional, and college-career needs of students (Carey & Martin, 2015). School counseling has been linked to increased college applications (Bryan, Moore-Thomas, Day-Vines, & Holcomb-McCoy, 2011), greater school bonding or connectedness (Bryan, Moore-Thomas, Gaenzle, Kim, Lin, & Na, 2012; Lapan, Wells, Petersen, & McCann, 2014; Lee & Smith-Adcock, 2005), better academic achievement (Carey & Martin, 2015), and higher school attendance (Carey & Martin, 2015). However, a need exists for more rigorous and policy-relevant research to demonstrate the effect of school counseling practices and programs on student academic, socio-emotional, and college-career outcomes (Bryan, Day-Vines, Holcomb-McCoy, & Moore-Thomas, 2010; Carey & Martin, 2015; Whiston, 2002).

National secondary datasets represent potential gold mines of data for researchers. Federal government agencies and private foundations fund these datasets and make them available to researchers to conduct policy research about a wide range of education; medical, socio-emotional, and mental health; cognitive and non-cognitive, and family and community constructs and issues related to children, youth, and adults. These national secondary datasets present a valuable source of data that could be used to conduct policy-relevant research on education and mental health issues related to school counseling as well as the effects of school counseling practices on student outcomes (Bryan, Day-Vines, Holcomb-McCoy, & Moore-Thomas, 2010; Carey & Martin, 2015). While a plethora of national secondary datasets exist that are designed for policy-relevant research related to K-12 and higher education, mental health, and public health, little use of these secondary datasets has been used in school counseling research. In 2010, in their article *Using National Education Longitudinal Datasets in School Counseling Research*, Bryan, Day-Vines, Holcomb-McCoy, and Moore-Thomas highlighted the low use of secondary education datasets by school counseling

J. Bryan (✉)
The Pennsylvania State University,
State College, PA, USA
e-mail: jabryan@psu.edu

J. Kim
Ball State University, Muncie, IN, USA
e-mail: jkim4@bsu.edu

Q. Shi
Loyola University Maryland, Baltimore, MD, USA
e-mail: qshi@loyola.edu

J.C. Carey et al. (eds.), *International Handbook for Policy Research on School-Based Counseling*,
DOI 10.1007/978-3-319-58179-8_11

researchers. At that time, few school counseling-related studies (i.e., Adams, Benshoff, & Harrington, 2007; Bryan, Holcomb-McCoy, Moore-Thomas, & Day-Vines, 2009; Bryan, Moore-Thomas, Day-Vines, Holcomb-McCoy, & Mitchell, 2009; Coker & Borders, 2001; Lee & Smith-Adcock, 2005; Suh, Suh, & Houston, 2007; Trusty, 2002; Trusty & Niles, 2003, 2004) had been conducted using the Department of Education's National Center of Education Statistics (NCES) datasets, such as the National Educational Longitudinal Survey 1988 (NELS:88) and the Educational Longitudinal Survey 2002 (ELS 2002). Since then, a few more studies related to school counseling have been conducted (i.e., Bryan, Day-Vines, Griffin, & Moore-Thomas, 2012; Bryan, Moore-Thomas, Day-Vines, & Holcomb-McCoy, 2011; Bryan, Moore-Thomas, Gaenzle, Kim, Lin, & Na, 2012; Cholewa, Burkhardt, & Hull, 2015). Yet, such secondary datasets like the ones developed by NCES could be used to answer a wider range of questions concerning the effects of school counselors and some school counseling practices on student outcomes. Indeed, the current relevance of school counseling is indicated by NCES' inclusion of a survey of school counselors in its latest study, the High School Longitudinal Study 2009 (HSLS 2009). An inclusion of a national survey of school counselors in the HSLS 2009 is a recognition of the critical role school counselors play in promoting students' high school, post-secondary, and early career decisions.

The purpose of this chapter is to increase school counseling researchers' knowledge regarding how to access and analyze national secondary datasets. Our aim is to promote the use of national secondary datasets to extend school counseling outcome research on what school counseling practices work or do not work and to help school counselors and other educators understand broadly the education and mental health characteristics of students, families, and communities that they serve. In this chapter we discuss the benefits and challenges of using these datasets, describe some of the existing datasets, delineate a research process to facilitate the process of using the datasets, and briefly explore some of the policy-relevant questions that could be answered with the datasets. We hope that readers will develop a clear understanding of the logistics and steps involved in conducting sound research studies with national secondary datasets.

The Benefits of Using Secondary Datasets

Research based on large national or international datasets often provides important insights for education policy-makers and decision-makers seeking to address students' academic, personal/social, and career issues. Further, they provide researchers with rich opportunities to examine academic gaps, inequitable educational opportunities, and factors related to education and counseling. Table 11.3 in the Appendix contains a list and descriptions of some of the datasets frequently used in school-based, youth, and mental health-related research. The American Psychological Association (APA) also provides a website with links to secondary datasets and data repositories suitable for educational, psychological, and youth-related research (see http://www.apa.org/research/responsible/data-links.aspx). These datasets are easily accessible to researchers who are interested in generalizing research findings to the larger target populations, exploring patterns among subpopulations in the representative data, and analyzing complex issues within multiple contexts and levels (Hahs-Vaugh, 2005; Osborne, 2011). Given the difficulties in collecting data from large samples due to cost and time, secondary datasets are also helpful for tenure-track and junior researchers allowing them to utilize data typically from larger samples to answer research questions related to counseling and educational issues (Osborn, 2011). Below, we expand on some of the benefits researchers derive from using national and international secondary datasets. Table 11.1 provides a summary of the benefits of using these datasets.

Table 11.1 Benefits and challenges of using existing secondary national and international datasets

Benefits	Cost-effectiveness	Save time and cost required to collect lots of data including standardized test scores and socio-emotional and behavior indicators Effective tool for junior faculty to conduct research given constrained pre-tenure time period
	Generalizability	The nationally representative sample enhances the generalizability of the target population
	Multiple data sources, contexts, and levels	Provide opportunities to investigate ecological and multiple factors with multiple data sources (e.g., standardized test scores, transcripts, and college enrollment), multiple individuals (e.g., students, administrators, parents, and school counselors), and multiple contexts (school, home, and community)
	Interdisciplinary studies	Allow the incorporation of different views from different fields such as education, sociology, and educational psychology that impact policy and practice
	Exploratory and comparative studies	Increase analysis options such as cross-sectional and longitudinal (or panel) comparison studies with national and international datasets
	Longitudinal investigation and causal inference	Allow the examination of causal relationships and longitudinal effects
	Advanced research design and methodology	Provide opportunities for developing or applying knowledge and skills that likely contribute to advance research design and methodology
Challenges	Methodological issues	Need to understand and develop knowledge of complex sampling design
	Fit between the data and research questions	Impossible to have control over data collection Difficulty in finding variables that are compatible with theoretical frameworks or constructs Difficulty in examining specific local or district issues
	Limiting validity and reliability	Single item or short scales that may reduce construct validity and internal reliability
	Nuances of datasets	Need to understand codebooks and manuals to address nuances of datasets such as response categories, skip patterns, and missing data
	Knowledge of the datasets	Need to invest time and energy to be knowledgeable about the datasets Learning statistical techniques from workshops, online trainings, and special seminars
	Advanced statistical analysis	Require knowledge and use of advanced statistical analysis Beneficial to have a research study team

Cost-Effectiveness

Secondary datasets offer researchers a cost-effective means of accessing large representative samples and multiple data sources such as standardized test scores, follow-up test scores, and data on students, parents, and school personnel like teachers and counselors and data on neighborhood and community factors (Hahs-Vaughn, 2007; Hofferth, 2005; Kluwin & Morris, 2006; Mueller & Hart, 2011; Nathans, Nimon, & Walker, 2013). Indeed, researchers likely save numerous time and cost in the data collection process due to public access of these already existing data (Kluwin & Morris, 2006). The availability of these datasets is important for junior faculty who are under pressure in the tenure and promotion process to publish substantive papers in a timely manner (Hofferth, 2005). For instance, NCES datasets such as the High School Longitudinal Study 2009 (HSLS 2009) provide students' standardized test scores and scores on a diverse range of socio-emotional and behavioral indicators that enable researchers to examine important relationships about student outcomes through various conceptual frameworks. The findings produced by these large-scale secondary data analyses may provide policy-makers and practitioners with valuable information to help improve student academic and mental health outcomes.

Generalizability

Many national and international secondary datasets design their data collection to produce nationally representative data from large samples. Nationally representative samples enhance the generalizability of findings from studies using these data (Kluwin & Morris, 2006; Nathans, Nimon & Walker, 2013; Strayhorn, 2009). Large samples that represent broad populations lead to greater precision in statistical estimation and increased generalizability (Hofferth, 2005). For instance, the Parent and Family Involvement in Education (PFI) survey from the National Household Education Surveys (NHES) developed by NCES addresses parents' and families' educational involvement, parents' postsecondary educational plans for their children, and factors related to parent educational participation and involvement (Herrold & O'Donnell, 2008). Using such data, researchers may draw generalizations about parents' involvement patterns and trends in relationship to their children's academic, socio-emotional, and college and career outcomes in the general population (Strayhorn, 2009).

Multiple Data Sources, Contexts, and Levels

Data on individuals in large national datasets are typically collected from multiple data sources. For example, in the ELS 2002 and HSLS 2009, data on students are collected from surveys of students themselves, parents, teachers and counselors, other school personnel, and school administrators, as well as directly from student records. These data provide researchers with information about students in multiple contexts such as the classroom, family, and even neighborhood contexts. As a result, researchers are able to use these data about classroom, school, family, and community characteristics to explore the influence of systemic or ecological factors on students' academic, socio-emotional, college, and career development (Mueller & Hart, 2011). For example, Espelage (2014) emphasizes the importance of using secondary datasets to examine bullying from an ecological perspective. Indeed, school counseling has increasingly emphasized an ecological framework, that is, the influence of student, family, school, and community characteristics and stakeholders and their interactions on students' lives and in addressing students' needs and problems (Bryan & Henry, 2012; McMahon, Mason, Daluga-Guenther, & Ruiz, 2014). Information about multiple contexts (e.g., home, school, and community environments), from multiple data sources (e.g., standardized test scores, transcripts, and postsecondary enrollment) and from multiple individuals (e.g., parents, teachers, school administrators, and students), allow researchers a unique opportunity to explore how ecological factors impact student outcomes. These rich data allow counseling researchers to test theories and models that can provide substantive information for counselors to guide the development of comprehensive and systemic interventions.

Interdisciplinary Studies

Secondary data may provide researchers in multiple disciplines with opportunities to explore educational or counseling topics from different research perspectives (Mueller & Hart, 2011). That is, national representative datasets allow interdisciplinary research from various counseling or educational disciplines, in which findings may provide valuable information to guide policy and practice in school-based counseling.

Exploratory and Comparative Studies

Numerous data analysis options exist with secondary datasets. The types of analyses vary with whether the data are cross sectional (e.g., National Survey of American Life) or longitudinal (or panel; e.g., NELS:88, ELS 2002, HSLS 2009) and whether it is structured to allow comparisons with other secondary datasets. See Table 11.3 in the Appendix to see which datasets are cross sectional or longitudinal in nature. Four types of comparison are possible: (a) cross-sectional comparisons in which one compares individuals and groups on data collected at one point, (b) longitudinal (or panel) comparisons in which one compares individuals and groups on data collected at more than one point in time (i.e., allows examination of changes over time or individual heterogeneity), (c) inter-cohort comparisons in which one compares individuals and groups across datasets on the same variables (e.g., NELS:88, ELS 2002, HSLS 2009), and (d) international comparisons in which one compares individuals and groups across countries (e.g., the Program for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS); Orletsky, Middleton, & Sloane, 2015, p. 316). National secondary datasets often enable researchers to compare issues in the USA to international contexts to provide insightful information on policy-relevant issues important to the worldwide community (Hahs-Vaughn, 2007; Wennberg, 2005). For instance, scholars can explore the academic performance of U.S. students to other countries using datasets such as National Assessment of Educational Progress (NAEP) or PISA.

Longitudinal Investigation and Casual Inference

Longitudinal data collected at multiple time points from the same samples provide opportunities to examine change over time as well as antecedent or mediator variables in statistical models with rigorous analysis. For instance, researchers can conduct longitudinal investigation on a college-going culture and identify precursors or other factors that predict later college enrollments in circumstance that cannot be manipulated experimentally (Grammer, Coffman, Ornstein, & Morrison, 2013). Due to the range of variables available in the dataset, researchers are also able to, in examining causal relationships, statistically control for confounding variables and selection bias that may affect results when using observational (nonexperimental) data (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007).

Advanced Research Design and Methodology

Secondary datasets provide opportunities for doctoral and junior faculty researchers to acquire advanced knowledge and skills in dealing with large samples and amounts of data and applying advanced statistical procedures (Hofferth, 2005). Researchers can increase their knowledge and skills regarding complex sampling design and weighting issues and the use of newly acquired knowledge of statistical analyses such as multilevel modeling (Stapleton & Thomas, 2008). Moreover, large longitudinal and representative samples enable scholars to use various statistical methodologies that may not be possible in the small sample sizes typically found with collection of primary data.

The Challenges of Using Secondary Datasets

Despite the many benefits, it is important that researchers understand and consider challenges to use secondary data. Below are detailed descriptions to be taken into consideration for using secondary datasets. Table 11.1 provides a summary of both the benefits and challenges of using these datasets.

Methodological Issues

Secondary data have methodological challenges that must be addressed by researchers who are concerned with generalizing findings to the intended population. Most samples in large national and international datasets are generally collected by cluster, stratified, or multistage sampling, rather than by random sampling (Bryan, Day-Vines, Holcomb-McCoy, & Moore-Thomas, 2010; Hahs-Vaughn, 2005, 2006). This type of complex sampling design creates statistical challenges such as larger standard errors and increased risk of Type I error, resulting in a need to use provided sample weights and consider the design effects. Therefore, researchers must take the time to understand complex sampling design and how to correct for design effects and apply sample weights (Hahs-Vaughn, 2007; Nathans, Nimon & Walker, 2013; Osborne, 2011). For instance, NCES data were collected using complex sampling designs, that is, multistage sampling including cluster and stratified sampling to ensure representation of the student population in the USA. This complex sampling design makes it important to employ appropriate analyses that produce accurate estimations of variances so as to avoid inaccurate results (Hahs-Vaughn, 2005, 2006, 2007). Further, to represent the intended target population, these data are often collected using oversampling procedures to increase the numbers of small groups in the population (e.g., minority groups). Hence, researchers should use the weights provided and proper weighting techniques (e.g., statistical software that allows application of weights) so as to arrive at accurate parameter estimates and findings that are generalizable to the population (Hahs-Vaughn, 2007; Orletsky, Middleton, & Sloane, 2015).

Fit Between the Data and Research Questions

Compared to primary data sources, secondary data were not designed for counselors only, so researchers have little control over data collection including the types of variables and questions that are explored (Bryan, Day-Vines, Holcomb-McCoy, & Moore-Thomas, 2010; Strayhorn, 2009). Therefore, researchers may have difficulty in finding specific variables in the secondary dataset that are compatible with their theoretical frameworks or constructs (Hofferth, 2005; Mueller & Hart, 2011). That is, the dataset may not include variables that fit the research question that a researcher wants to examine. For instance, when a researcher is interested in students' sense of purpose, it may be hard to find appropriate items for identifying or operationalizing the construct. Also, nationally representative samples may not reflect specific issues in particular contexts or populations (Strayhorn, 2009). For instance, when researchers are interested in educational issues in school districts and the effects of local policies or neighborhood factors on academic achievement, it may be difficult to use national secondary data to gain specific information at the district or state levels (Warren, 2015).

Limiting Validity and Reliability

Another challenge with datasets is that often only a single item or a few items are available to measure a construct. For example, researchers often measure concepts or constructs relevant to school counseling issues, such as social capital, student-counselor contact, and parent empowerment, with single items and short scales. This may mean that a theoretical construct is measured incompletely (Grammer, Coffman, Ornstein, & Morrison, 2013). Thus, these single items and short scales may decrease the construct validity and internal reliability of the measures used and may impact the degree of precision and error with which researchers measure the construct they want to measure (Hofferth, 2005; Wennberg, 2005).

Nuances of Datasets

Researchers need to be familiar with codebooks and technical manuals to understand and address the nuances of datasets such as handling missing data and coding procedures (Hahs-Vaughn, 2007). For instance, the variables of interests may need to be recoded as alphanumeric scaled and Likert-type items (Hahs-Vaughn, 2007). Also, it is important to understand complex skip patterns and patterns of missing data so as to establish statistical plans to deal with them in data analysis.

Knowledge on the Datasets

Although the secondary datasets are a cost-effective tool, it is still necessary to invest time and energy to understand the data collection process, documentation, and structure of the data files in order to utilize the datasets appropriately and accurately (Hofferth, 2005; Strayhorn, 2009). Some training may be necessary to learn strategies and techniques for conducting secondary data analysis with existing datasets. NCES, American Institute of Research (AIR), and American Educational Research Association (AERA) offer workshops, online trainings, and special seminars to equip education researchers with the skills to access and use datasets (Bryan et al., 2010; Hahs-Vaughn, 2007).

Advanced Statistical Analysis

The complex dataset may make it difficult to conduct a simple study. The nature of the datasets, especially those comprising longitudinal and multiple samples, better lend themselves to advanced statistical analyses, especially in instances when scholars are interested in long-term follow-up of participants and complex contextual factors (Hofferth, 2005). For example, researchers may need to use multilevel modeling analysis in order to answer research questions on the roles of specific levels of the school environment in students' academic, social, and career outcomes. Collaborating with a research team and acquiring funding are beneficial when undertaking intensive studies with these datasets (Grammer, Coffman, Ornstein, & Morrison, 2013).

Identifying and Using the Datasets: The Research Process

A large number of datasets are available that provide wonderful opportunities for researchers to conduct substantive theory-driven and model testing studies that contribute to existing knowledge. Table 11.3 in the Appendix provides a list and description of some of the more common datasets used in counseling, education, and mental health-related fields. To help researchers take advantage of these opportunities, in this section we describe what we consider to be a step-by-step research process to conducting studies with these datasets.

The Six-Step Research Process for Using and Evaluating Secondary Datasets

Bryan, Day-Vines, Holcomb-McCoy, and Moore-Thomas (2010) and Hofferth (2005) provided a six-step research process and a number of useful questions to guide research with national secondary datasets. Here, we update and expand Bryan et al.'s (2010) and Hofferth's (2005) discussions under the following revised framework: (a) gaining access to the dataset, (b) getting to know the data and evaluating its suitability for the study, (c) preparing the data for analysis, (d) conducting appropriate data analyses, (e) interpreting results and examining implications including policy implications, and (f) considering and describing the limitations of the data. Figure 11.1 shows the six steps in the research process, while Table 11.2 delineates the important tasks and subtasks at each stage of the research process using secondary datasets.

Step 1: Gaining Access to the Dataset

Accessing the Data

The first step in the research process is for researchers to identify and gain access to the appropriate dataset(s). Table 11.3 in the Appendix describes a wide range of secondary datasets with information about their purpose, the nature of the data, and where information about the data

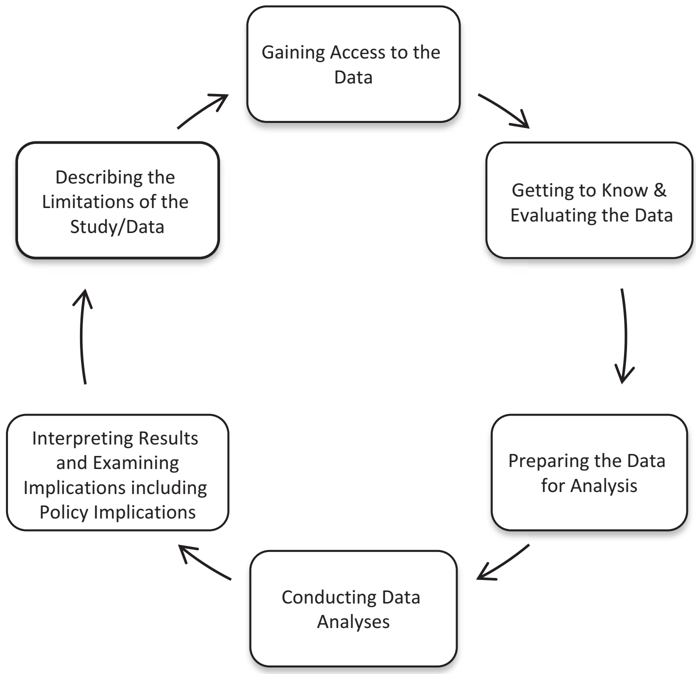

Fig. 11.1 Six-step research process for using and evaluating secondary datasets

may be obtained. Data repositories such as the National Center of Education Statistics (NCES), the Institute of Social Research (ISR) including the Inter-university Consortium Center of Political and Social Research (ICPSR), and the National Data Archive on Child Abuse and Neglect (NDACAN) at Cornell University house a large number of national and international secondary datasets that allow researchers to examine educational, mental, and public health and sociopolitical questions and issues pertaining to children and their caregivers such as parents, school counselors, and other school personnel (See Table 11.3). Many of these datasets are public use datasets and available through the websites of these organizations. Some datasets contain restricted data such as identifiers of schools and zip codes that could lead to researchers identifying participants with some effort (Strayhorn, 2009). When datasets contain these types of identifying data, the owners of the datasets often restrict access to these datasets or to the identifying variables. To gain access to restricted data, researchers must apply for restricted use licenses. In most cases, restricted data must be housed in a secure location at the researchers' institution and on computers without access to the Internet and available only to researchers named on the license.

Gathering Information

Researchers typically start with tentative questions based on their research interests, and the more information they gather about the dataset(s) will help them to determine its suitability for their research in the early stages of the research process. The information-gathering step is integral to understanding the dataset. The first steps in gathering information entail going to the websites and manual to read about the datasets and examine the surveys used to collect the data and the codebooks to see what variables are found in the dataset. Secondly, researchers should read journal articles about studies utilizing the dataset. In addition, numerous articles exist on how to use secondary datasets in addition to this chapter. See our reference list for a wide variety of articles on using secondary datasets. Although

Table 11.2 The process of using large secondary datasets

Stages of the research process	Steps/tasks at each stage of the process	Important subtasks at each step
Step 1: Gaining access to the dataset	Accessing the data	Accessing public use data of interest (usually available on a website) Applying for access to restricted data (Applying for a license)
	Gathering information	Gathering information with your research interests and questions in mind Reading manuals, survey questionnaires, and literature that describe and explain the methodological features of the studies used to collect the data Reading journals and research literature that used the dataset Reading articles on how to use secondary datasets
	Finding training opportunities	Researching opportunities for online training and/or face-to-face training Checking the due dates for training applications
Step 2: Getting to know the data and evaluating its utility/suitability to your study	Becoming familiar with the dataset	Determining whether dataset has variables related to your research questions Understanding items and what they measure (how variables are operationalized) Becoming familiar with previous research and how previous researchers measured similar constructs
	Determining appropriate research questions and the suitability of the dataset for the study	Identifying and refining appropriate questions Evaluating suitability of data for answering these questions Considering the policy relevance of the research questions
	Developing the conceptual framework (iterative process)	Examining frameworks used by previous researchers who used the dataset Examining researcher positionality Challenging conceptual frameworks for study of marginalized groups
	Challenging the conceptual frameworks through which you examine marginalized groups	Examining research positionality on marginalized groups Being cautious presenting ethnic variation within minority groups

(continued)

Table 11.2 (continued)

Stages of the research process	Steps/tasks at each stage of the process	Important subtasks at each step
Step 3: Preparing the data for analysis	Creating a usable dataset	Setting up the data appropriately Making sure one is familiar with previous studies/data analyses using the dataset Keeping a copy of the working file so can start over from scratch if necessary
	Using sampling weights and strata and cluster variables to control for complex data effects	Choosing correct weight for analysis Choosing computer software vs manual application of weight to adjust for design effects Recognizing that using software to adjust for design effects is more accurate
	Selecting appropriate statistical software	Choosing appropriate statistical software that allows application of weights, strata, and cluster variables
	Handling missing data and determining the analytic sample	Using best practices for missing data management (e.g., multiple imputation, maximum likelihood estimation)
	Choosing items and creating composites	Choosing items based on theory Creating composites using factor analysis and/or theory Comparing use of the same or similar items in other studies
Step 4: Conducting appropriate data analyses	Building analysis from foundation up to more complex analyses	Using multiple methods to answer the research questions (to tell the story/create a fuller picture) Beginning with single-level and univariate analyses and then move to multivariate and multilevel analyses
	Using more advanced statistical methods	Examining individual heterogeneity (within variance) using multilevel approaches (e.g., HLM)
	Replicate your studies	Replicating or encouraging replication of your studies Comparing results of other similar studies using similar methods
Step 5: Interpreting the results and examining the implications including policy implications	Writing policy implications	Making connections between findings and policy implications in practical, relevant, and concrete ways Thinking about policy relevance from the outset of research
Step 6: Considering and describing the limitations of study	Limitations of secondary data sources	Understanding the secondary data as proxies for the construct researchers intend to measure Unable to capture the quality of the relationships in the items Not overstretching from the results Not overgeneralizing to those not represented by the sample

researchers are eager to get their hands on the datasets and get started on analyses, it would be a huge mistake to begin data analysis without gathering information about the methodology behind the data and reading the research literature on studies completed using the same or similar data (Lauritsen, 2015).

Finding Training Opportunities

While organizations that own national datasets and their affiliates often conduct trainings for some datasets, some trainings are currently conducted online. For example, prior to 2013, NCES conducted annual onsite trainings for researchers. Now researchers can engage in self-directed trainings for the many NCES datasets using their Distance Learning Dataset Training (DLDT) website found at https://nces.ed.gov/training/datauser/, which comprises modules on the NCES datasets. These modules introduce you to each dataset, its purpose, information on the data collection, sampling design, sampling weights, and data analysis considerations. Trainings on some of the most recent NCES datasets are also conducted at a number of national conferences each year including the American Educational Research Association (AERA) national conference. In addition, AERA conducts an annual Institute on Statistical Analysis to promote researchers' use of current NCES datasets to examine policy-related questions of interest in a particular area (e.g., postsecondary transitions, mathematics education). Applications to the Statistical Institute are usually due in January each year. The Association for Institutional Research (AIR) also provides online and face-to-face education on NCES datasets including an annual NCES Data Institute, cosponsored by NCES. Applications for the Data Institute are typically due in February each year. The Inter-university Consortium for Political and Social Research (ICPSR) also offers a number of courses each year in its summer school that focus on selected datasets such as a four-week summer workshop on Quantitative Analysis of Crime and Criminal Justice sponsored by the Bureau of Justice Statistics (BJS).

Step 2: Getting to Know the Data and Evaluating Its Utility/Suitability to Your Study

Becoming Familiar with the Dataset

Once you have gained access to the data, perhaps the most important step of all is getting to know the data. Researchers should begin by reading the manuals and reports that describe the data. For example, NCES provides a detailed manual describing background, instrumentation, sample design, coding systems, and sample weights for all of its surveys (see http://nces.ed.gov/surveys/). It is imperative that researchers understand how survey items were worded and structured, what participants' responses to the survey items of interest mean, what the response options were, and who actually answered the question (Wells, 2016; Wennberg, 2005). Sometimes the item is not really measuring what it appears to be measuring. Taking the time to explore these measurement issues will help researchers determine the extent to which the items may be used to measure the variables in their proposed study and address their research questions. Further, researchers will find it helpful to become familiar with previous research and how other researchers have used the same items in published studies, especially studies that used the same dataset or similar ones (Hofferth, 2005; Wells, 2016).

Determining Appropriate Research Questions and the Suitability of the Dataset for the Study

Appropriate research questions emerge when researchers engage in an iterative process of immersing themselves in the theoretical and empirical literature and closely examining the manuals and survey questions of one or more secondary datasets in order to determine their suitability for their research interests. If the dataset is not suitable for answering the research questions, researchers may need to reconceptualize the study and/or seek new data (Hofferth, 2005). While researchers should pursue the research questions that capture their interest, as they decide on their questions, they should also consider how their questions and study could contribute to policy related to students, families, and school-based counseling.

Developing the Conceptual (or Theoretical) Framework

As with any other research, researchers who are conducting studies with secondary datasets should be guided by strong conceptual/theoretical frameworks. A strong conceptual/theoretical framework improves researchers' decisions about what variables to use, what constructs to measure, and what items define these constructs, the type of appropriate data analyses, and interpretation of results. Conceptual or theoretical frameworks emerge from the conceptual and empirical literature on these ideas, concepts, and variables as well as from researchers' own personal experiences. All of this together forms a "tentative theory" (Maxwell, 2012, p. 36) of what you think is going on between and among the concepts, ideas, factors, and variables of interest. Research without a conceptual framework is simply data mining (Zhang, 2010). In large datasets, researchers will inevitably find relationships between variables, but a lack of consistent theoretical/conceptual underpinnings that explain the relationships among the variables will undermine the credibility of their study. Taking time to be immersed in the literature and develop a conceptual framework will improve one's research goals, research questions, justification for the study, methodological decisions, selection and validity of the measures, and interpretation of the results.

In the extant research, counseling researchers have used a variety of conceptual frameworks to develop studies with secondary datasets. These frameworks include school bonding (Bryan, Moore-Thomas, Gaenzle, Kim, Lin, & Na, 2012; Lee & Smith-Adcock, 2005), social capital (Bryan, Moore-Thomas, Day-Vines, & Holcomb-McCoy, 2011; Croninger & Lee, 2001; Perna & Titus, 2005), parent empowerment (Kim & Bryan, 2017), discipline disproportionality (Bryan, Day-Vines, Griffin, & Moore-Thomas, 2012), and college opportunity structure (Engberg & Gilbert, 2014). These conceptual frameworks helped researchers bring new insights to the problems studied.

Challenging the Conceptual Frameworks Through which You Examine Marginalized Groups

In conducting research with large datasets, we caution researchers to examine their views or positionality about the research participants or the problem they are studying. Researchers' positionality refers to researchers' worldview and beliefs that affect the stance they take to the research and the problem they are studying (Foote & Bartell, 2011; Milner, 2007). Their positionality influences the language and narratives they use to describe marginalized groups (e.g., female, racial or ethnic minority, immigrant, poor, urban, or rural students and families). Researchers must be careful not to perpetuate the negative attributes and stereotypes of people of color and other marginalized groups (Milner, 2007). For example, often education researchers define children as "at risk" or "problematic" without examining the complexities of the problem or realities of the participants, or the profound impact of the labels on children (Swadner, 1990). Although researcher positionality is a term used mostly in the qualitative research literature, it is important for all researchers to be critical of their use of language and how they are presenting their findings.

Relatedly, it is important to be cautious about presenting marginalized groups as monolithic groups. The tendency for researchers to ignore ethnic variation within minority groups may hinder a deeper understanding of educational processes among racial/ethnic groups. For example, studying Black children as one racial group fails to take into account the difference in history, cultures, social, and family experiences among Black children (Griffith, Neighbors, & Johnson, 2009; O'Connor, Lewis, & Mueller, 2007). Yet, the experiences of Caribbean, African, African-American, and Hispanic Black children differ in many ways.

Step 3: Preparing the Data for Analysis

Creating a Useable Dataset

Preparing a useable dataset for analysis from a national secondary dataset takes an investment of time and thought. Researchers must first create a working dataset that contains the variables and sample they want to use in their study and that is ready for analysis (Willms, 2011). This working file should include any sample weights, cluster, strata, and identification variables. The weight, cluster, and strata variables are particularly important for use in data analyses with complex samples. Researchers should make sure that all missing values and label values are coded correctly. It is wise to keep a copy of this working dataset in case the need arises to revert to it to start analyses over from the beginning.

Using Sampling Weights and Strata and Cluster Variables to Control for Complex Sample Design Effects

Data in most large national datasets are collected from complex samples meaning that the samples were selected using stratified, cluster, or multistage sampling or a combination of both. Therefore, the data cannot be treated like a simple random sample. A large number of articles discuss the importance of using sampling weights in the statistical analysis of complex samples and specific procedures for doing so (Hahs-Vaughn 2005, 2006, 2007; Hahs-Vaughn, McWayne, Bulotsky-Shearer, Wen, & Faria, 2011a, 2011b; Osborne, 2011; Wells, Lynch, & Seifert, 2011). When sample weights are applied, the sample size increases to represent the population (Osborne, 2011). For example, the 21,444 students in the HSLS 2009 are representative of over three million students, and when the student sample weight is applied, the sample size in analyses is extremely large. As a result of these large sample sizes, data analyses that do not account for the complex design effects will produce smaller standard errors or increased Type I error (Bryan et al., 2010; Hahs-Vaughn, 2005, 2006). Therefore, researchers must make a correction for these artificially small standard errors that increase the likelihood of significant results (Type I error). To correct for the complex sample design effects, researchers must choose and apply the appropriate sampling weight for their analysis. When the software allows it, they should also apply the strata and cluster (primary sampling unit) variables. This process requires selecting the appropriate statistical software, choosing the appropriate weight for the analytic samples, and selecting the strata and cluster variables provided in the datasets.

Selecting Appropriate Statistical Software

Numerous statistical software packages now accommodate complex samples, such as SPSS, SAS, Stata, MPlus, HLM, and Latent Gold (Hahs-Vaughn, McWayne, Bulotsky-Shearer, Wen, & Faria, 2011a). These software packages allow researchers to specify the weight, strata variable, and cluster variable (also called primary sampling unit or PSU in some datasets). These software allow researchers to conduct analyses that automatically adjust for the complex design effects and result in more accurate analyses. However, when one does not have statistical software for complex survey data or when one only has the sampling weight (e.g., strata and cluster variables are withheld in some restricted datasets such as HSLS 2009), researchers may make corrections manually (Bryan et al., 2010). Previous practices for manually adjusting the sample weight include scaling (or re-normalizing) the weight to adjust the sample size (Hahs-Vaughn, 2005, 2006; Osborne, 2011). For example, most NCES datasets include two types of design effects. Researchers can use these NCES-derived design effects: (1) the root design effect (*deft*) to adjust the standard errors of test statistics or (2) the average design effects (*deff*) to create a new weight, that is, to renormalize the weight. Researchers who wish to manually adjust and apply the weight may use either one of these design effects to renormalize the appropriate weight. These calculations are described in detail elsewhere (Bryan et al., 2010; Hahs-Vaughn, 2005, 2006).

Handling Missing Data and Determining the Analytic Sample

The analytic sample is the sample on which you will conduct your analyses on and is arrived at by cleaning the data, selecting all the relevant data, and handling missing data (Bryan et al., 2010). Too much missing data will limit the generalizability of a study's findings; hence, researchers should make decisions about which procedure they will use to deal with missing data (Bryan et al., 2010; Hahs-Vaughn, 2007; Wells, 2016). Some traditional methods include dropping observations using listwise or pairwise deletion and mean substitution (i.e., replacing missing values on a variable with the mean of the variable). However, these methods can lead to reduced sample size and limited variability and may affect your findings; therefore, it is better to use model-based methods to handle missing data such as multiple imputation and maximum likelihood estimation procedures. See the following work (Allison, 2003; Baraldi & Enders, 2010; Peugh & Enders, 2004; Schlomer, Bauman, & Card, 2010) for an in-depth discussion of strategies for managing missing data.

Choosing Items and Creating Composites

Choosing items and creating composite variables to represent variables should not be done haphazardly, but should be guided by researchers' conceptual framework and, wherever possible, a factor analytic approach (e.g., principal components analysis, principal factor analysis, and confirmatory factor analysis). For example, if a researcher is measuring school bonding or school connectedness, both his/her conceptual framework and a factor analysis will be helpful in determining which items to select, how to name the factors or components, and the validity (i.e., whether a measure or factor actually measures what it purports to measure) of your measures. Many national datasets contain categorical measures; therefore, researchers may need to use a nonlinear factor analysis method for nominal or ordinal items such as nonlinear (or categorical) principal component analysis (also known as CATPCA). For examples of CATPCA with datasets, see Hahs-Vaughn (2017) and Kim and Bryan (2017). When researchers need to select a single item or a few items as a proxy for a construct they wish to measure, it is equally important that they use theory as well as a comparison of how other researchers use these same items or similar ones in studies that measure the same construct or variable.

Step 4: Conducting Appropriate Data Analyses

After selecting appropriate software; determining the appropriate weights, cluster, and strata variables; and conducting a missing data analysis to arrive at their analytic sample, researchers are be ready to analyze the data using statistical methods that are suitable for their research questions and the type of data (Bryan et al., 2010; Wells, 2016). Researchers should not be discouraged if they do not have the statistical knowledge necessary to conduct all of the analyses we highlight in this section. Statistics are mere tools to answer questions, and this knowledge can be required through taking online and face-to-face courses and workshops as well as through reading of some of the user-friendly texts and articles we recommend later in this section. Researchers should also collaborate with colleagues who are knowledgeable about the statistical methods they desire to use to answer their research question. Information on webinars and short courses and workshops can be found at websites such as the Analysis Factor (http://www.theanalysisfactor.com/about/) and Statistical Horizons (http://statisticalhorizons.com) and at a number of university summer programs in quantitative methods, such as ICPSR's Summer Program in Quantitative Methods of Social Research (http://www.icpsr.umich.edu/icpsrweb/sumprog/index.jsp) and the Odum Institute (http://www.odum.unc.edu/odum/contentSubpage.jsp?nodeid=21).

Building Analyses Foundation Up to More Complex Analyses

Researchers should use multiple methods to tell a story, to answer the research question, and to paint a full picture of the phenomenon and relationships under study (Wells, Lynch & Seiffert, 2011). Rather than yielding to the temptation to jump straight to the more complex analyses (e.g., factor analysis, multiple regression, structural equation modeling, multilevel modeling, and latent class analysis), it is important to build the house from the bottom up (Willms, 2011). Starting with simple methods such as descriptive analyses (e.g., mean, median, mode, skewness, standard deviation, frequencies, and proportions) and correlational analyses allows one to disaggregate the data and provide greater insight about outliers, differences among the subgroups in the sample, and linear and nonlinear associations among the variables. These analyses help researchers to tell the story and to paint the big picture and, ultimately, allow researchers to better understand and explain results from their more advanced analyses, often uncovering any underlying patterns and meanings in the more complex results (Bryan et al., 2010).

Using More Advanced Statistical Methods

In particular, national secondary datasets are most suitable to more advanced statistical techniques (Bryan et al., 2010; Wells, 2016). The most common analyses with secondary datasets appear to be multivariate analyses, like multiple linear regression and logistic and ordinal regression analyses. Multiple regression can be a useful lens for examining findings from national datasets (Nathans, Nimon, & Walker, 2013). However, many datasets comprise data collected from groups of individuals (e.g., students, parents, and teachers) who are clustered or nested within higher level units (e.g., classrooms, schools, colleges, neighborhoods, organizations, and countries) or they comprise repeated observations on individuals over time (i.e., observations nested in individuals). These multilevel data violate the independence assumption in many traditional statistical procedures such as multiple regression because individuals in the same cluster (e.g., students in the same classroom) are more alike (or homogeneous) and their scores are dependent. This lack of independence may result in increased Type I error rates and incorrect results when using statistical procedures based on the independence assumption (Peugh, 2010; Zhang, 2010). Multilevel modeling (also called hierarchical linear modeling [HLM]) allows researchers to take advantage of these nested data to examine change within persons (i.e., individual differences or heterogeneity) or within units (Cheslock & Aguilar, 2011; Lynch, 2012; Peugh, 2010; Zhang, 2010). Counseling researchers should explore the benefits of more advanced statistical procedures such as structural equation modeling (Byrne, 2016), multilevel modeling or hierarchical linear modeling (Lynch, 2012; O'Connell & McCoach, 2008; Peugh, 2010; Snijders & Bosker, 2012), and latent class analysis (LCA; Collins & Lanza, 2013; Lanza & Cooper, 2016; Lanza & Rhoades, 2013) where appropriate for answering their research questions.

Replication of Studies

We encourage researchers who use secondary datasets to replicate their studies with other datasets to build a knowledge base and test previous findings. The fact that many secondary datasets collect the same or similar data on participants over a period of time allows for replication to see if findings are consistent. For example, Dumais (2009, 2008) compared teenagers in the NELS:88 and ELS:2002 cohorts and found consistent patterns in academic attitudes, extracurricular participation, and math achievement among 12th graders. The lack of replication of studies in school-based education and counseling often makes it difficult for researchers to make strong conclusions (Makel & Plucker, 2014). Replication builds on previous research while at the same time working to establish a body of credible knowledge about a particular phenomenon (Nathans, Nimons, & Walker, 2013, p. 26–27). Indeed, the replication of studies with these datasets provides a more credible knowledge base from which to make policy and practice recommendations (Makel & Plucker, 2014).

Step 5: Interpreting the Results and Examining the Implications Including Policy Implications

Often counseling researchers fail to provide policy implications from their research. Like many researchers, they identify implications for practice and future research, but fail to go a step further to identify policy-relevant conclusions or to present them in a persuasive manner (Glover, 2002). However, when counseling researchers conduct studies with national secondary datasets, they must recognize that these datasets are often constructed to guide policy on pertinent educational and mental health problems facing schools, families, communities, and governments. Hence, policy-makers are interested in the policy implications from these studies (Bryan et al., 2010; St. John, 2004). Therefore, it is important that researchers think carefully about the policy implications of their studies with these datasets rather than providing broad recommendations and vague conclusions that could fit almost any topic on school-based counseling (e.g., the school should increase the number of school counselors..., or school counselors need more training to…, or school counselors should consider these findings when…). In presenting the findings and implications, it is important to ask oneself whether a policy-maker (the audience) would find this paper credible, be clear about what action needs to be taken, and see the recommendations as practical, relevant, and concrete (Glover, 2002; Wilcox & Hirschfield, 2007). To be successful in producing policy-relevant research and implications, policy should not be an afterthought at the end of the study. Indeed, it is important that counseling researchers think about the policy relevance and implications of their research from the outset as they develop their research plan and design.

Step 6: Considering and Describing the Limitations of the Study

Researchers must be transparent about limitations that exist with the use of data from national and international secondary datasets. Although secondary datasets have many strengths, they also bring some limitations as reflected in the challenges discussed earlier in this chapter. First, researchers are constrained by the fact that the data is collected by someone else and they often have to use items to measure constructs which the items may not have been intended to measure (Bryan et al., 2010; Kluwin & Morris, 2006; Wennberg, 2005). This results in researchers utilizing items as proxies for complex constructs, for example, school bonding (Bryan, Moore-Thomas, Gaenzle, Kim, Lin, & Na, 2012; Lee & Smith-Adcock, 2005), social capital (Bryan, Moore-Thomas, Day-Vines, & Holcomb-McCoy, 2011; Croninger & Lee, 2001; Perna & Titus, 2005), and parent empowerment (Kim & Bryan, 2017). These measurement issues highlight the importance of a strong conceptual or theoretical framework to guide researchers' operationalization of variables as well as the importance of comparing items used to measure similar or related constructs across studies that utilized secondary datasets (Bryan et al., 2010; Wells, 2016). Relationships, such as counselor-student relationships and students' contact with school counselors and other helping professionals in schools, are of great interest to school counseling researchers. While these data allow researchers to examine the effect of students' or parents' contact with these professionals, they do not reveal the quality or extent of these interactions (Bryan et al., 2010). Moreover, in some cases, researchers are limited to one item in measuring student-counselor contact (see Bryan, Holcomb-McCoy, Moore-Thomas, & Day-Vines, 2009; Bryan, Moore-Thomas, Day-Vines, & Holcomb-McCoy, 2011) or parent-counselor contact (see Kim & Bryan, 2017). Not only does the use of one or two items to measure a variable affect the construct validity and reliability of the variable; it also limits the conclusions and rich interpretations researchers may make. In general, researchers must be careful about overstretching their conclusions and should openly discuss these limitations (Wells, Lynch & Seiffert, 2011).

Potential Research Questions Related to Policy Issues on School-Based Counseling

To date, counseling researchers have examined a number of problems and issues pertinent to schools with national secondary datasets. These issues include predictors of students' contact with the school counselor for college information (Bryan, Holcomb-McCoy, Moore-Thomas, & Day-Vines, 2009) and for counseling services

(Bryan, Moore-Thomas, Day-Vines, Holcomb-McCoy, & Mitchell, 2009), predictors of high school dropout (Suh, Suh, & Houston, 2007), effects of parent empowerment on academic achievement (Kim & Bryan, 2017), effects of school counselor-student contact on college application rates (Bryan, Moore-Thomas, Day-Vines, & Holcomb-McCoy, 2011), effects of school bonding on academic achievement (Bryan, Moore-Thomas, Gaenzle, Kim, Lin, & Na, 2012), predictors of school counselors' influence on underrepresented students' thoughts about post-secondary education (Cholewa, Burkhardt, & Hull, 2015), and how counseling opportunity structure varies among schools and affects their college enrollment rates (Engberg & Gilbert, 2014). However, greater understanding of these issues as well as a number of other important policy-relevant issues could be developed through thoughtful secondary data research using the six-step research process in this chapter.

Notably, policy-makers are concerned with critical issues such as closing academic gaps, reducing educational inequity, and promoting college and career readiness in American schools. Potential research could provide valuable information for policy-makers to make decisions about what educational programs and strategies are effective to promote student academic, social/emotional, and career/college development. Below, we briefly discuss potential areas of school counseling research that have direct policy implications for school-based counseling practice.

Relationship between School Counseling Practices and Student Outcomes

School counseling research could examine factors related to school counseling practices/interventions and students' academic and college outcomes. Given the importance of the roles school counselors play in promoting academic and college outcomes, school counselors and policy-makers are interested in the effectiveness of school counseling practices and interventions to best serve all students. Some national datasets provide variables explicitly suited to examine connections between counseling-related activities and student outcomes. Specifically, the HSLS:2009 includes school counselor questionnaires about school counseling practices and interventions related to college and career readiness. Using this dataset, school counseling researchers could examine counseling factors that may influence college readiness (e.g., school counselor-student contact, GPA, SAT, and taking advanced courses), college choices (e.g., application to 2-year vs. 4-year colleges or to nonselective vs. selective universities), and college enrollment. Further, the research could identify how school counseling practices may affect traditionally underrepresented students' outcomes (e.g., immigrant students, students from lower SES families, first-generation college students, English language learners, and students with disabilities). Such research could provide specific information on which practices and interventions might be most beneficial to students and how counselors can better serve and advocate for students who are underrepresented in schools.

Noncognitive Variables that Influence Students' Academic and College Outcomes

The American School Counselor Association (ASCA, 2014) has recently called attention to the noncognitive factors or the academic mindsets that are integral to students' performance and college readiness. "*Academic mindsets* are beliefs, attitudes, or ways of perceiving oneself in relation to learning and intellectual work that promote academic performance" (Nagaoka, Farrington, Roderick, Allensworth, Keyes, Johnson, & Beechum, 2013, p. 49). Researchers from the University of Chicago Consortium on Chicago School Research has highlighted the importance of noncognitive factors to academic success (Farrington, Roderick, Allensworth, Nagaoka, Keyes, Johnson, & Beechum, 2012; Nagaoka, Farrington, Roderick, Allensworth, Keyes, Johnson, & Beechum, 2013). Thus, future school counseling research could explore the influence of specific variables (e.g., self-concept, sense of belonging, perseverance, self-efficacy, sense of purpose) on student academic and college outcomes. For instance, Bryan, Moore-Thomas, Day-Vines, and Holcomb-McCoy (2011) examined the relationship between school bonding and high school students' academic outcomes using educational longitudinal dataset (ELS:02). Future investigations could replicate the school bonding research regarding students' college choices and enrollment. Moreover, school counseling research could examine the relationship of academic mindsets (e.g., sense of belonging, self-efficacy) and college readiness to college enrollment and retention.

Students' College and Career Pathways

President Obama's administration called for educators to ensure that all students are well prepared for college and careers (Bryan, Young, Griffin, & Henry, 2016). School counselors are at an optimal position to help address students' college access and attainment as well as their career preparation. Indeed, school counselors play crucial roles in planning and preparing students for postsecondary education, including 2-year and 4-year colleges and technical and vocational schools (Bryan, Young, Griffin, & Henry, 2016). However, in the current school counseling literature, very little school counseling research exists on counseling factors that may influence students' postsecondary application and enrollment decisions, their choice of major in college, and their future career pathways. School counseling research could examine postsecondary variables (e.g., postsecondary aspirations, career aspirations, application to various types of postsecondary institutions, work and career experiences, college enrollment, and degree attainment) that influence students' college and career pathways. For instance, given the importance of STEM enrollment for the country's continued prosperity, more research is warranted about the factors that could potentially influence students' STEM-related college/vocational/career major choices. Important implications for school counselors' practice and programming and policies related to students' college and career development may emerge from this research.

School Violence and Bullying

School violence and bullying always have been issues of concern in U.S. schools. About a third of middle and high school students are physically bullied and over half are verbally bullied (U.S. Department of Education, 2010). However, most of the articles focus on offering recommendations, strategies, and interventions for school counselors, parents, teachers, communities, and legal systems (Allen, 2010). Very little research exists that provides national findings about what school counseling services could help address school violence and bullying issues. National data allow researchers to investigate how ecological factors, including school counseling practices, may contribute to school violence and bullying. Further, school counseling researchers could examine the relationships and effects of bully victimization and school violence to and on students' academic, behavioral, and college outcomes across multiple contexts (e.g., family, school, and community). For instance, NCES datasets (e.g., Early Childhood Longitudinal Study (ECLS), ELS 2002, and HSLS 2009) include student items that assess school violence, bullying victimization, problem behaviors, peer influence, and school safety (Espelage, 2014, 2015). Moreover, the School Crime and Safety (SCS) survey would allow researchers to explore how promotive, protective, and risk factors might be conducive to creating positive or negative environments that discourage school violence and bullying (Espelage, 2014, 2015). Further, these datasets can be used to explore how perceptions of teachers, parents, and school adults may mediate school violence and student outcomes (Espelage, 2014, 2015). Such investigations may provide information for school counselors that inform how the design of prevention efforts for reducing school violence and bullying.

Conclusion

Our aim has been to provide step-by step logistics to help researchers conduct school counseling research that impact policy decisions utilizing national secondary datasets. We offer a six-step research process that school counseling researchers may follow to conduct research to examine the effects of school counseling and school practices and programs on student academic, socio-emotional, and college-career outcomes. While they may present several challenges in terms of the learning curve needed to understand methodological and statistical issues, these datasets provide valuable and cost-effective opportunities to conduct rigorous, generalizable, longitudinal, and casual studies to advance knowledge. The six-step research process may act as a map which guides researchers to their final research goals by delineating tasks at each step of the process and informing how to complete them. In the six-step process, we emphasize the importance of a theoretical framework, researcher positionality, and policy implications that researchers should consider and challenge throughout the entire process. If researchers desire to conduct school counseling research with national secondary datasets, they could examine school counseling and other educational factors related to academic or opportunity gaps and educational inequities, and academic, college/career, and mental health outcomes, which all have implications for developing effective school counseling interventions and practices.

Appendix

Table 11.3 A summary of national datasets suitable for counseling research

Dataset	Brief description	Cross-sectional/longitudinal (#of waves)	Participants in the study	Country data collected in	Where data housed	Websites
Educational longitudinal studies						
High School Longitudinal Study of 2009 (HSLS: 09)	Nationally representative, longitudinal study of 23,000+ 9th graders from 944 schools in 2009; students followed throughout secondary and postsecondary years	Longitudinal (three waves: base year 2009, first follow-up 2012, second follow-up planned for 2016)	Students, parents, math and science teachers, school administrators, and school counselors	USA	NCES	https://nces.ed.gov/surveys/hsls09/index.asp
Education Longitudinal Study of 2002 (ELS: 2002)	Nationally representative, longitudinal study that followed a sample of 15,000 tenth graders in 2002 and 12th graders in 2004 throughout their secondary and postsecondary years.	Longitudinal (four waves: base year 2002, first follow-up 2004, second follow-up 2006, third follow-up 2012)	Students, parents, math and English teachers, and school administrators	USA	NCES	http://nces.ed.gov/surveys/els2002/
National Education Longitudinal Study of 1988 (NELS:88)	Nationally representative study of 24,599 eighth graders from 1035 schools on the following topics: school, work, and home experiences, educational resources and support, the role in education of their parents and peers, neighborhood characteristics, educational and occupational aspirations, and other student perceptions.	Longitudinal; five waves: Based year: 1988 First follow-up: 1990 Second follow-up: 1992 Third follow-up: 1994 Fourth follow-up: 2000	Students, teachers, parents, and school administrators	USA	NCES	https://nces.ed.gov/surveys/nels88/
Early Childhood Longitudinal Program (ECLS)–Kindergarten Class of 1998–1999 (ECLS-K)	Nationally representative study of 22,666 children from the beginning of their kindergarten through middle school (5–13 years old). Focuses on children's status at entry to school, their transition into school, and their progression through 8th grade	Longitudinal; five waves: Base year: 1998–1999 First follow-up: 1999–2000 Second follow-up: 2002 Third follow-up: 2004 Fourth follow-up: 2007	Students, parents, teachers and school administrators	USA	NCES	https://nces.ed.gov/ecls/kindergarten.asp

Early Childhood Longitudinal Study 2010–2011 (ECLS-K: 2011)	Nationally representative study of 20,000 kindergarteners through fifth grade (5–10 years old). Focuses on descriptive information on children's status at entry to school, their transition into school, and their progression through the elementary grades	Longitudinal; five waves: Base year: 2010–2011 First follow-up: 2011–2012 Second follow-up: 2012–2013 Third follow-up: 2014 Fourth follow-up: 2015 Fifth follow-up: planned for 2016	Children, their families, teachers, schools, and care providers	USA	NCES	https://nces.ed.gov/ecls/kindergarten2011.asp
National Longitudinal Transition Study-2 (NLTS-2)	Nationally representative study of 11,500 students (13–16 years old) in 2000 as they moved from secondary school into adulthood	Longitudinal; five waves: Wave1: 2001 Wave2: 2003 Wave3: 2005 Wave4: 2007 Wave5: 2009	Students, parents, and teachers and principals	USA	IES	http://www.nlts2.org and http://ies.ed.gov/ncser/projects/nlts2/index.asp
National Longitudinal Transition Study-2012 (NLTS: 2012)	Nationally representative study of 15,000 students as they transition to young adults. The goal is to better understand how school programs are supporting all students in this transition period, including those with special needs.	Longitudinal; two waves: Wave I: 2012–2013 Wave II: 2014	Students, parents, teachers, and school administrators	USA	IES	http://ies.ed.gov/ncee/nlts/index.asp
National Household Education Surveys (NHES)	Nationally representative study of all ages (early childhood through school age to adulthood). Focuses on descriptive information about educational activities and condition of education in the USA	Cross sectional; data collected in 1991, 1995, 1999, 2001, 2003, and 2005	Students and parents	USA	NCES	http://nces.ed.gov/nhes/index.asp
NHES Parent and Family Involvement in Education (PFI-NHES): 2007, 2012	PFI-NHES is one of the major repeating surveys on school-age children in the NHES program. PFI focuses on parent and family involvement in children's education	Data collected in 1996, 1999, 2003, 2007, and 2012	Parents or guardians	USA	NCES	https://nces.ed.gov/nhes/surveytopics_school.asp
Longitudinal studies of children and youth						
The National Longitudinal Study of Adolescent to Adult Health (ADD Health)	The ADD Health followed a nationally representative sample of adolescents in grades 7–12 until they reached young adulthood, providing data on respondents' social, economic, psychological, and physical well-being	Longitudinal, five waves: Wave I: 1994–1995 Wave II: 1996 Wave III: 2001–2002 Wave IV: 2008–2009 Wave V: planned for 2016–2018	Adolescent students, parents, school administrators	USA	University of North Carolina	http://www.cpc.unc.edu/projects/addhealth

(continued)

Table 11.3 (continued)

Dataset	Brief description	Cross-sectional/longitudinal (#of waves)	Participants in the study	Country data collected in	Where data housed	Websites
Fragile Families and Child-Wellbeing Study (FFCWS)	FFCWS is designed to address (1) what are the conditions and capabilities of unmarried parents, especially fathers, (2) what is the nature of the relationships between unmarried parents, (3) how do children born into these families fare, and (4) how do policies and environmental conditions affect families and children	Five waves of publicly available data including baseline (at birth) and year one, three, five, and nine	Follows a cohort of nearly 5000 children born in large US cities between 1998 and 2000. Both mothers and fathers are interviewed	USA	Joint effort by Princeton University and Columbia University	http://www.fragilefamilies.princeton.edu
Monitoring the Future	8th, 10th, and 12th grade students respond to drug use and demographic questions, as well as to additional questions on a variety of subjects, including attitudes toward religion, parental influences, changing roles of women, educational aspirations, self-esteem, exposure to sex and drug education, and violence and crime – both in and out of school	Ongoing yearly study of the behaviors, attitudes, and values of American secondary school students, college students, and young adults	Each year, a total of approximately 50,000 8th, 10th, and 12th grade students are surveyed (12th graders since 1975 and 8th and 10th graders since 1991)	USA	Institute of Social Research at the University of Michigan	https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/35 and http://www.monitoringthefuture.org
Longitudinal Study of American Youth (LSAY)	A national study to allow the nation to understand the thinking and the life experiences of Generation X	Longitudinal; two cohorts: 1987–1994, merged cohort: 2007 New 7th grade cohort: 2015	Students, parents, science and math teachers, and principals	USA	University of Michigan	http://lsay.org
Longitudinal Study of Young People in England (LSYPE)	LSYPE was set up to gather evidence about the transitions young people make from secondary and tertiary education or training to economic roles in early adulthood, enhance the ability to monitor and evaluate the effects of existing policy and provide a strong information base for future policy development; and contextualize the implementation of new policies in terms of young people's current lives	Longitudinal; seven waves	Students and parents	England	Department of Education in UK	https://www.education.gov.uk/ilsype/workspaces/public/wiki/LSYPE

Longitudinal Studies of Child Abuse and Neglect (LONGSCAN)	LONGSCAN is funded by the National Center on Child Abuse and Neglect to permit study of critical issues of child maltreatment. Assessments can be used alone or combined for pooled analysis.	Longitudinal. Starting in 1990, data were collected every 2 years from children (and their parents and teachers) at ages 4, 6, 8, 12, 14, 16, and 18 years	Students, parents, and teachers	USA	University of North Carolina Chapel Hill	http://www.unc.edu/depts/sph/longscan/ Can also be found at the National Data Archive on Child Abuse and Neglect (NDACAN) http://www.ndacan.cornell.edu
Current Population Survey Civic Engagement Supplement	Provide information on communication with others, interaction with public institutions and private enterprises, forming positive relationships with others, participation in groups, extent of political action, and frequency of gaining news and information from media sources	November 2008, 2009, 2010, 2011, 2013, and 2014	Adult participants	USA	US Department of Commerce Bureau of the Census; Sponsored by Bureau of Labor Statistics and Corporation for National and Community Service (CNCS)	https://catalog.data.gov/dataset/current-population-survey-civic-engagement-supplement
The Commission on Youth Voting and Civic Knowledge Youth Post-Election Survey 2012	A study of 4483 participants aged 18–24 about their political participation and their educational experiences	Interviews began the day after the 2012 presidential election and continued on for 6 weeks after the election	Youth aged 18–24	USA	Turfs University Center for Information and Research on Civic Learning and Engagement (CIRCLE)	http://www.icpsr.umich.edu/icpsrweb/civicleads/studies/35012#datasetsSection
National datasets on mental health, violence, and delinquency						
National Crime Victimization Survey (NCVS)	As the nation's primary source of information on criminal victimization, the NCVS provides the largest national forum for victims to describe the impact of crime and characteristics of violent offenders	Cross sectional; ongoing yearly data available from 1973 to 2014	12 years + in each sampled household	USA	University of Michigan and Bureau of Justice Statistics	http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/95 and http://www.bjs.gov/index.cfm?ty=dcdetail&iid=245

(continued)

Table 11.3 (continued)

Dataset	Brief description	Cross-sectional/longitudinal (#of waves)	Participants in the study	Country data collected in	Where data housed	Websites
School Crime Supplement to the National Crime Victimization Survey (SCS/NCVS)	Using a national survey with about 6500 students (12–18 years old), SCS collects information about victimization, crime, and safety at school in US public and private elementary, middle, and high schools	The SCS was conducted in 1989, 1995, 1999, 2001, 2003, 2005, 2007, 2009, 2011, and 2013	Students aged 12–18	USA	NCES	https://nces.ed.gov/programs/crime/student_data.asp
2009 National Survey on Drug Use and Health (NSDUH)	With annual nation and state-wide interviews with 70,000 participants, NSDUH provides information on the use of tobacco, alcohol, illicit drugs, and mental health in the USA	Annual survey since 1988	Youth of Ages 12 and above	USA	Substance Abuse and Mental Health Services Administration (SAMHSA)	SAMHSA https://nsduhweb.rti.org/respweb/homepage.cfm
National Survey of American Life (NSAL)	The primary goal of the NSAL was to gather data about the physical, emotional, mental, structural, and economic conditions of Black Americans at the beginning of the new century	Cross sectional; 1year	African-American, Afro-Caribbean, and non-Hispanic white adults, age 18+ residing in households in the coterminous USA. Exclusions include institutionalized persons, those living on military bases, and non-English speakers	USA	University of Michigan	http://www.icpsr.umich.edu/icpsrweb/CPES/

National Latino and Asian American Study (NLAAS)	The NLAAS provides national information on the similarities and differences in mental illness and service use of Latinos and Asian Americans.	Cross sectional	Latino and Asian-American adults, age 18+ residing in households in the coterminous USA, Alaska, and Hawaii. Exclusions include institutionalized persons and those living on military bases.	USA	Center for Multicultural Mental Health Research	http://www.multiculturalmentalhealth.org/nlaas.asp and http://www.icpsr.umich.edu/icpsrweb/CPES/files/nlaas
Pathways to Desistance study	The Pathways to Desistance study is a multisite, longitudinal study of serious adolescent offenders as they transition from adolescence into early adulthood	Longitudinal	Adjudicated youths between 14 and 18 years old	USA	University of Michigan	http://www.icpsr.umich.edu/icpsrweb/NAHDAP/series/260
International education datasets						
The Trends in International Mathematics and Science Study (TIMSS)	TIMSS provides reliable and timely data on the mathematics and science achievement of US students compared to that of students in other countries. Data have been collected from 4th and 8th graders since 1995 every 4 years, generally	Cross sectional	Students (math and science)	More than 60 countries and other education systems	Boston College	http://timssandpirls.bc.edu
Program for International Student Assessment (PISA)	International assessment that measures 15-year-old students' reading, mathematics, and science literacy every 3 years	Cross sectional	Students (major domain of study rotates between mathematics, science, and reading in each cycle)	More than 70 countries and educational jurisdictions	OECD	http://www.oecd.org/pisa/aboutpisa/

References

Adams, J. R., Benshoff, J. M., & Harrington, S. Y. (2007). An examination of referrals to the school counselor by race, gender, and family structure. *Professional School Counseling, 10*, 389–398.

Allen, K. (2010). A bullying intervention system: Reducing risk and creating support for aggressive students. *Preventing School Failure, 54*(3), 199–209. doi:10.1080/10459880903496289.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545–557.

American School Counselor Association. (2014). *Mindsets and behaviors for student success: K–12 college-and career-readiness standards for every student*. Alexandria, VA: Author.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*(1), 5–37.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routledge.

Bryan, J., Day-Vines, N., Griffin, D., & Moore-Thomas, C. (2012). The disproportionality dilemma: Patterns of teacher referrals to school counselors for disruptive behavior. *Journal of Counseling and Development, 90,* 177–190.

Bryan, J., Day-Vines, N. L., Holcomb-McCoy, C., & Moore-Thomas, C. (2010). Using national education longitudinal data sets in school counseling research. *Counselor Education and Supervision, 49*(4), 266–279.

Bryan, J., & Henry, L. (2012). A model for building school–family–community partnerships: Principles and process. *Journal of Counseling and Development, 90*, 408–420. doi:10.1002/j.1556-6676.2012.00052.x.

Bryan, J., Holcomb-McCoy, C., Moore-Thomas, C., & Day-Vines, N. (2009). Who sees the school counselor for college information? A national study. *Professional School Counseling, 12,* 280–291.

Bryan, J., Moore-Thomas, C., Day-Vines, N., & Holcomb-McCoy, C. (2011). School counselors as social capital: The effects of high school college counseling on college application rates. *Journal of Counseling and Development, 89*, 190–199.

Bryan, J., Moore-Thomas, C., Day-Vines, N., Holcomb-McCoy, C., & Mitchell, N. M. (2009). Characteristics of students who receive school counseling services: Implications for practice and research. *Journal of School Counseling, 7*(21), 1–29. Retrieved from http://jsc.montana.edu/articles/v7n21.pdf.

Bryan, J., Moore-Thomas, C., Gaenzle, S., Kim, J., Lin, C., & Na, G. (2012). The effects of school bonding on high school seniors' academic achievement. *Journal of Counseling and Development, 90,* 467–480.

Bryan, J., Young, A., Griffin, D., & Henry, L. (2016). Preparing students for higher education: How school counselors can foster college readiness and access. In J. L. DeVitis & P. Sasso (Eds.), *Higher education and society* (pp. 149–172). New York: Peter Lang.

Carey, J. C., & Martin, I. (2015). *A review of the major school counseling policy studies in the United States: 2000–2014*. Amherst, MA: Center for School Counseling Outcome Research and Evaluation. Retrieved from http://www.counseling.org/docs/default-source/resources---school-counselors/carey-and-martin-2015-policy-research-monograph.pdf?sfvrsn=8.

Cheslock, J. J., & Rios-Aguilar, C. (2011). Multilevel analysis in higher education research: A multidisciplinary approach. In J. Smart & M. Paulsen (Eds.), *Higher education: Handbook of theory and research* (Vol. 26, pp. 85–123). The Netherlands: Springer.

Cholewa, B., Burkhardt, C. K., & Hull, M. F. (2015). Are school counselors impacting underrepresented students' thinking about postsecondary education? A nationally representative study. *Professional School Counseling, 19*(1), 144–154.

Coker, J. K., & Borders, L. D. (2001). An analysis of environmental and social factors affecting adolescent problem drinking. *Journal of Counseling & Development, 79*, 200–208.

Collins, L. M., & Lanza, S. T. (2013). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (2nd ed.). New York: John Wiley & Sons.

Croninger, R. G., & Lee, V. E. (2001). Social capital and dropping out of high school: Benefits to at-risk students of teachers' support and guidance. *Teachers College Record, 103*, 548–581.

Dumais, S. A. (2009). The academic attitudes of American teenagers, 1990–2002: Cohort and gender effects on math achievement. *Social Science Research*, *38*(4), 767–780.

Dumais, S. A. (2008). Cohort and gender differences in extracurricular participation: The relationship between activities, math achievement, and college expectations. *Sociological Spectrum, 29*(1), 72–100.

Engberg, M. E., & Gilbert, A. J. (2014). The counseling opportunity structure: Examining correlates of four-year college going rates. *Research in Higher Education, 55*(3), 219–244.

Espelage, D. (2015). Data needs for emerging research issues in bully and violence prevention: Strengths and limitations of the National Center for Educational Statistics data sets. *AERA Open*, *1*(3), doi: http://dx.doi.org/10.1177/2332858415604147.

Espelage, D. L. (2014). *Using NCES surveys to understand school violence and bullying.* Paper Prepared for the National Academy of Education's Workshop to Examine Current and Potential Uses of NCES Longitudinal Surveys by the Education Research Community, Washington, DC. Retrieved from http://www.naeducation.org/cs/groups/naedsite/documents/webpage/naed_160696.pdf.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago:

University of Chicago Consortium on Chicago School Research. Retrieved from https://consortium.uchicago.edu/sites/default/files/publications/Noncognitive Report.pdf.

Foote, M. Q., & Bartell, T. G. (2011). Pathways to equity in mathematics education: How life experiences impact researcher positionality. *Educational Studies in Mathematics, 78*(1), 45–68.

Glover, D. (2002). *What makes a good "policy paper"? Ten examples* (no. sp200205t1). Economy and environment program for Southeast Asia (EEPSEA). Retrieved from https://idl-bnc.idrc.ca/dspace/bitstream/10625/27168/1/118102.pdf.

Grammer, J. K., Coffman, J. L., Ornstein, P. A., & Morrison, F. J. (2013). Change over time: Conducting longitudinal studies of children's cognitive development. *Journal of Cognition and Development, 14*(4), 515–528. Retrieved from http://curran.web.unc.edu/files/2015/03/BainterCurran2015.pdf.

Griffith, D. M., Neighbors, H. W., & Johnson, J. (2009). Using national data sets to improve the health and mental health of Black Americans: Challenges and opportunities. *Cultural Diversity and Ethnic Minority Psychology, 15*(1), 86.

Hahs-Vaughn, D. L. (2017). *Applied multivariate statistical concepts*. Abingdon, UK/Oxon/New York: Routledge.

Hahs-Vaughn, D. L. (2007). Using NCES national datasets for evaluation of postsecondary issues. *Assessment & Evaluation in Higher Education, 32*(3), 239–254. http://education.ufl.edu/educational-research/files/2011/01/Hahs-Vaughn_2007_Using-NCES-datasets.pdf.

Hahs-Vaughn, D. L. (2006). Analysis of data from complex samples. *International Journal of Research & Method in Education, 29*(2), 165–183.

Hahs-Vaughn, D. L. (2005). A primer for using and understanding weights with national datasets. *Journal of Experimental Education, 73*(3), 221–248. Retrieved from http://education.ufl.edu/educational-research/files/2011/01/Hahs-Vaughn_2005_Using-weights-with-national-datasets.pdf.

Hahs-Vaughn, D. L., McWayne, C. M., Bulotsky-Shearer, R. J., Wen, X., & Faria, A. M. (2011b). Complex sample data recommendations and troubleshooting. *Evaluation Review,* 0193841X11412070. Retrieved from https://www.researchgate.net/profile/Rebecca_Bulotsky-Shearer/publication/51245439_Complex_sample_data_recommendations_and_troubleshooting/links/53da4e0b0cf2e38c633676b6.pdf.

Hahs-Vaughn, D. L., McWayne, C. M., Bulotsky-Shearer, R. J., Wen, X., & Faria, A. M. (2011a). Methodological considerations in using complex survey data: An applied example with the head start family and child experiences survey. *Evaluation Review, 35*(3), 269–303. Retrieved from https://www.researchgate.net/profile/Rebecca_Bulotsky-Shearer/publication/51245439_Complex_sample_data_recommendations_and_troubleshooting/links/53da4e0b0cf2e38c633676b6.pdf.

Herrold, K., & O'Donnell, K. (2008). *Parent and Family Involvement in Education, 2006–07 school year, from the National Household Education Surveys Program of 2007 (NCES 2008–050)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Hofferth, S. L. (2005). Secondary data analysis in family research. *Journal of Marriage and Family, 67*(4), 891–907.

Kim, J., & Bryan, J. (2017). A first step to a conceptual framework of parent empowerment: Exploring relationships between parent empowerment and academic performance in a national sample. *Journal of Counseling and Development, 95*(2), 168–179.

Kluwin, T. N., & Morris, C. S. (2006). Lost in a giant database: The potentials and pitfalls of secondary analysis for deaf education. *American Annals of the Deaf, 151*(2), 121–128.

Lanza, S. T., & Cooper, B. R. (2016). Latent class analysis for developmental research. *Child Development Perspectives, 10*(1), 59–64. doi:10.1111/cdep.12163.

Lanza, S. T., & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science, 14*(2), 157–168. http://doi.org/10.1007/s11121-011-0201-1.

Lapan, R. T., Wells, R., Petersen, J., & McCann, L. A. (2014). Stand tall to protect students: School counselors strengthening school connectedness. *Journal of Counseling & Development, 92*(3), 304–315.

Lauritsen, J. L. (2015). The devil is in the details: Crime and victimization research with the National Crime Victimization Survey. In M. D. Maltz & S. K. Rice (Eds.), *Envisioning criminology: Researchers on research as a process of discovery* (pp. 191–197). Cham, Switzerland: Springer International Publishing.

Lee, S., & Smith-Adcock, S. (2005). A model of girls' school delinquency: School bonding and reputation. *Professional School Counseling, 9*(1), 78–87.

Lynch, M. F. (2012). Using multilevel modeling in counseling research. *Measurement and Evaluation in Counseling and Development, 45*(3), 211–224.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43*, 304–316. 0013189X14545513.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (3rd ed.). Thousand Oaks, CA: Sage.

McMahon, H. G., Mason, E., Daluga-Guenther, N., & Ruiz, A. (2014). An ecological model of professional school counseling. *Journal of Counseling & Development, 92*(4), 459–471.

Milner, H. R. (2007). Race, culture, and researcher positionality: Working through dangers seen, unseen, and unforeseen. *Educational Researcher, 36*(7), 388–400.

Mueller, C. E., & Hart, C. O. (2011). Effective use of secondary data analysis in gifted education research: Opportunities and challenges. *Gifted Children, 4*(2), 3. Retrieved from http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1048&context=giftedchildren.

Nagaoka, J., Farrington, C. A., Roderick, M., Allensworth, E., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2013). Readiness for college: The role of Noncognitive factors and context. *Voices in Urban Education, 38*, 45–52.

Nathans, L. L., Nimon, K., & Walker, D. A. (2013). Use of the multiple lens approach to multiple regression findings with a national dataset. *Multiple Linear Regression Viewpoints, 39*(1), 25–33. Retrieved from http://mlrv.ua.edu/2013/vol39_1/Nathans%20et%20al..pdf.

O'Connor, C., Lewis, A., & Mueller, J. (2007). Researching "Black" educational experiences and outcomes: Theoretical and methodological considerations. *Educational Researcher, 36*, 541–552.

Orletsky, D., Middleton, J. A., & Sloane, F. (2015). A review of three large-scale datasets critiquing item design, data collection, and the usefulness of claims. In J. A. Middleton, J. Cai, & S. Hwang (Eds), *Large-Scale Studies in Mathematics Education* (pp. 311–334). Research in Mathematics Education Series. Springer International Publishing. Springer.

Osborne, J. (2011). Best practices in using large, complex samples: The importance of using appropriate weights and design effect compensation. *Practical Assessment Research and Evaluation, 16*, 1–7. Retrieved from http://pareonline.net/pdf/v16n12.pdf.

O'Connell, A. A., & McCoach, D. B. (2008). *Multilevel modeling of educational data*. Charlotte, NC: Information Age Publishing.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85–112.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525–556.

Perna, L. W., & Titus, M. A. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *Journal of Higher Education, 76*, 485–518.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. Washington, DC: American Educational Research Association. Retrieved from http://documents.library.nsf.gov/edocs/LB1028.E77-2007-PDF-Estimating-Causal-effects-using-experimental-and-observational-designs.pdf.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology, 57*, 1–10.

Snijders, T. A., & Bosker, T. A. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Los Angeles: Thousand Oaks: Sage.

Stapleton, L. M., & Thomas, S. L. (2008). The use of national datasets for teaching and research: Source and issues. In A. O'Connell & B. McCoach (Eds.), *Multilevel modeling of educational* (pp. 11–59). Charlotte, NC: Information Age Publishing.

Strayhorn, T. L. (2009). Accessing and analyzing national databases. In T. J. Kowalski & T. J. Lasley II (Eds.), *Handbook of data-based decision making in education* (pp. 105–122). New York: Routledge.

St. John, E. P. (2004). Public policy and research on college access: Lessons learned from NCES research and state evaluation studies. In E. P. S. John (Ed.), *Public policy and college access: Investigating the federal and state roles in equalizing postsecondary opportunity, readings on equal education* (Vol. 19, pp. 181–196). New York: AMS Press.

Suh, S., Suh, J., & Houston, I. (2007). Predictors of categorical at-risk high school dropouts. *Journal of Counseling and Development, 85*(2), 196.

Swadener, E. B. (1990). Children and families "at risk:" etiology, critique, and alternative paradigms. *The Journal of Educational Foundations, 4*(4), 17.

Trusty, J. (2002). African Americans' educational expectations: Longitudinal causal models for women and men. *Journal of Counseling & Development, 80*, 332–345.

Trusty, J., & Niles, S. G. (2003). High-school math courses and completion of the bachelor's degree. *Professional School Counseling, 7*, 99–107.

Trusty, J., & Niles, S. G. (2004). Realized potential or lost talent: High school variables and bachelor's degree completion. *The Career Development Quarterly, 53*, 2–15.

United States Department of Education. (2010). Anti-bullying policies: Examples of provision in state laws. Key policy letter from the Education Secretary and Deputy Secretary. Retrieved from http://www2.ed.gov/policy/gen/guid/secletter/101215/html.

Warren, J. R. (2015). The Future of NCES's Longitudinal Student Surveys. *AERA Open, 1*(2), 2332858415587910. Retrieved from http://ero.sagepub.com/content/1/2/2332858415587910.full OR http://ero.sagepub.com/content/spero/1/2/2332858415587910.full.pdf.

Wells, R. (2016). Large national datasets. In F. K. Stage & K. Manning (Eds.), *Research in the college context: Approaches and methods* (2nd ed., p. 106-?). New York: Routledge.

Wells, R. S., Lynch, C. M., & Seifert, T. A. (2011). Methodological options and their implications: An example using secondary data to analyze Latino educational expectations. *Research in Higher Education, 52*(7), 693–716. Retrieved from https://www.researchgate.net/profile/Ryan_Wells2/publication/225768412_Methodological_Options_and_their_Implications_An_Example_Using_Secondary_Data_to_Analyze_Latino_Educational_Expectations/links/0f317535c5ac351646000000.pdf.

Wennberg, K. (2005). Entrepreneurship research through databases: Measurement and design issues. *New England Journal of Entrepreneurship, 8*(2), 3. Retrieved from http://digitalcommons.sacredheart.edu/cgi/viewcontent.cgi?article=1148&context=neje.

Whiston, S. C. (2002). Response to the past, present, and future of school counseling: Raising some issues. *Professional School Counseling, 5*, 148–155.

Wilcox, A., & Hirschfield, A. (2007). *A framework for deriving policy implications from research.* University of Huddersfield: Applied Criminology Centre. Retrieved from http://www.scopic.ac.uk/documents/Huddersfield_1.pdf

Willms, J. D. (2011). Managing and using secondary data sets with multidisciplinary research teams. In K. H. Trzesniewski, M. Brent Donnellan, & R. E. Lucas (Eds.), *Secondary data analysis: An introduction for psychologists* (pp. 27–38). Washington, DC: American Psychological Association. x, 245 pp.http://dx.doi.org/10.1037/12350-002.

Zhang, L. (2010). The use of panel data models in higher education policy studies. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 25, pp. 307–349). Dordrecht, The Netherlands: Springer.