Designing Mixed Methods Analysis

6.1 Identifying Analysis Strategies

Mixed methods is an emergent research methodology that advances systematic and intentional integration, or "mixing," of qualitative and quantitative data within a single inquiry. When qualitative research or quantitative research limits or hinders your full understanding of a problem, this integration allows a more complete investigation than separate qualitative and quantitative studies. To be successful, you must be knowledgeable and confident in conducting both quantitative and qualitative analyses techniques. Alternatively, collaboration with others who have complementary methods skills can be a successful strategy. You must also know how to mix and even embed these analyses, so that the findings from each component allow you to combine inferences coming from both sets of analyses into one coherent whole (Onwuegbuzie et al., 2009; Tashakkori & Teddlie, 2003). By having a positive attitude toward both qualitative and quantitative methods, researchers are well placed to use qualitative research to inform the quantitative portion of research studies, and vice versa (Onwuegbuzie & Leech, 2005).

Using mixed methods analysis allows you to create a more complex design that seeks more comprehensive understanding of the study focus. When you mix, you get more than just a qualitative and quantitative study; you get a much richer view of the problem under investigation. You collect, analyze, and integrate both quantitative and qualitative data in a social problem under inquiry to address your study focus. This section considers how results can be merged, or integrated, for comparison; how qualitative data explain the quantitative results; or how quantitative results can help explain the qualitative results. Furthermore, what we may learn from qualitative or quantitative analysis can help guide us into the other type of data in ways that may not have been initially apparent to ask potentially new and important questions.

In this chapter, we show you how to use Dedoose to help you discover and explore patterns in your mixed methods data. Using the different tools available in the application, you can drill down into your data to better understand the rich qualitative stories that live beneath the surface.

Assuming that your research problem or question warrants a mixed methods study, we think the core characteristics of a well-designed mixed methods study must include the following:

1. Framing the inquiry using a strong theoretical framework

2. Designing both qualitative and separate quantitative components

133

3. Designing how to mix these qualitative and quantitative components—for example, either nested, sequential, or concurrent mixed methods designs, as discussed in Chapter 1, which combines the strengths, and overcomes the weaknesses, of a single method design

4. Using systematic and rigorous data gathering and analysis techniques appropriate to the chosen framework

5. Gathering both qualitative and quantitative data

6. Integrating the data during gathering, analysis, and/or discussion

7. Presenting findings drawing on both the qualitative and quantitative components

To help you with your analysis strategies, first list your quantitative and qualitative data sources. Having a clear idea of the data you will be dealing with and what you hope to accomplish as you enter your analytic process will help frame the overall picture of how things will unfold. Then think about the steps you will take to prepare and conduct your analysis: preparing your data, preliminary analysis, and steps to conduct and complete the analysis. While you are thinking about this, consider the skills, time, and resources you have available, and really challenge yourself to be realistic!

> ### BRUCE'S TIP #22
> Create Diagrams About Your Research
>
> It can be very helpful to draw a diagram of your research design (how you plan to mix your studies), including data gathering, theoretical framework, and analysis strategies. Make sure to keep key research questions and goals in mind throughout the process. *Remember to include the date and save each diagram as your study progresses.* As you find your way through your analysis, it can be very helpful to review this developing thinking.

Regardless of when you gather and analyze your data, Table 6.1 shows some starting points for you to think about as you move into your analysis.

When you have analyzed your data from both the qualitative and quantitative segments of your study, you must integrate the data in an appropriate way to fit with your study design. Integrating your data will allow you to maximize the strengths of each approach, while minimizing weaknesses. Effective integration points can occur during data collection, data coding and analysis, and drawing

	Qualitative Data	Quantitative Data
Table 6.1 Working With Your Data		
Preparing your data	Organizing your data, document management, preparing data for import into data application	Coding your data and assigning numeric values where necessary, preparing data for analysis in data application
Exploring	Reading data, memoing, coding	Descriptive analysis, looking for trends and distributions
Analyzing	Coding data and assigning labels, looking for related themes, use of data application	Using appropriate statistical tests, using statistical analytic features in Dedoose where applicable or other software, recording of confidence intervals

meanings or interpreting your data. Successfully integrating your data is how you are able to draw comprehensive conclusions from your work.

At this point, you are probably thinking: "How do I do this?" "Where do I start?" The following discussion shows you a variety of ways to approach these challenges. For now, be thoughtful about your goals, be flexible in thinking how to integrate and analyze your data, and remain focused realistically on the arguments you hope to communicate and can support.

To help you continue your developing thinking about mixed methods, here are some common analytic strategies for the integration of qualitative and quantitative data (Bryman, 2006; Caracelli & Greene, 1993; Onwuegbuzie & Teddlie, 2003):

- *Embedding the Data*

  Using this strategy, here you have to make some choices. Which data set do you consider to be the primary source? Once you have made this decision, the second set of data can be embedded in the first one. For example, the primary data in a project may relate to quantitative data gathered during a clinical trial. You may also have gathered qualitative narrative responses to enhance your data set. These narrative responses can be embedded in your primary data set.

- *Connecting the Data*

  When you are connecting your data, you analyze one set of data and you use your results to guide subsequent data gathering. In this way, you make a connection between the two data sets, but you are not directly comparing the results. This strategy can be used in two-phase projects where data are collected and analyzed sequentially, for example, an action research approach.

- *Data Reduction*

  Here, you reduce the dimensionality of the qualitative data (e.g., via memoing and exploratory thematic analysis) and quantitative data (e.g., via descriptive statistics, exploratory factor analysis, cluster analysis, and transforming data from sets of items to scale scores and from continuous variables to categorical variables).

- *Data Transformation, Consolidation, or Merging*

  Often researchers transform one data set so that it can be compared with the other data set. This conversion or transformation of one data type into the other allows you to then analyze both together:

  o Qualitative data are converted into numeric codes (which can be done using code weights in Dedoose) that can be included with quantitative data in statistical analyses.

  o Quantitative data are transformed into narrative data and included with qualitative data in thematic or pattern analysis.

  o Qualitative and quantitative data are jointly reviewed and consolidated into numerical codes or narrative for purposes of further analysis.

  For example, you might assign numeric codes to the narrative data collected in a qualitative study to enable comparison with the quantitative results. This joint review of both data types allows you to create new or consolidated variables or data sets, which can be expressed in either quantitative or qualitative form. You can then use these consolidated variables or data sets for further analyses.

- *Extreme Case Analysis*

  Identifying extreme, or deviant cases, is a sampling strategy that occurs within the context of and in conjunction with other sampling strategies. "Extreme cases" are unusual cases in the study or cases that are considered outliers. They are identified after some portion of data collection and analysis has been completed. These cases appear to be the "exception to the rule" emerging from the analysis. Once identified, these cases can be investigated through further data gathering or analysis to extend your understanding of any initial explanation for the extreme cases. By seeking out extreme or deviant cases, you can develop a richer, more in-depth understanding of a phenomenon and further strengthen your account of your research and what you are learning.

Good mixed methods studies have distinct, and well-developed, qualitative and quantitative components that are integrated or connected in some

intentional way with a sound rationale for doing so. Each strand has its own questions, data, analysis, and interpretations. Data gathering for both components is rigorous using advanced analytic techniques. Meaningful inferences are made from the results of each strand, and procedures promoting trustworthiness of the work are reported (e.g., member checks, triangulation, threats to internal validity, etc.; Creswell & Tashakkori, 2007). Trustworthy mixed methods research integrates or links the strands and describes how this integration has occurred, so it is clear how the conclusions drawn from the two strands provide a fuller understanding of the phenomenon under study. You might integrate your work by comparing, contrasting, building on, or embedding one type of conclusion with the other (Creswell & Tashakkori, 2007; O'Cathain, Murphy, & Nicholl, 2008a).

Having discussed what makes a good mixed methods study and how you might go about mixing your data, it is now time to think about using Dedoose to help you manage your data and carry out your analysis. Basically, the ways in which any REDA assists mixed methods analysis (other than general strategies for the qualitative component) are by

a. allowing contents of a range of different sources to be brought together (using coding) to facilitate complementary analysis;

b. facilitating importing quantitative data, for combination with qualitative data, usually to facilitate comparative analyses based on values of those quantitative variables;

c. identifying and exploring patterns that may draw on qualitative and/or quantitative aspects of the database; and

d. allowing for export of coding information for each case (or source, depending on the structure of the software) as a case by variable table, which can then be imported into statistical software for further processing (Bazeley, 2017).

6.2 Using Descriptors

Using descriptors in your analysis can be of great value, but remember, you are not required to use them in your research. While people use different terms for *descriptors*, this term is preferred in Dedoose because they are essentially information that "describes" a research participant or groups of participants and helps the researcher to distinguish between them. For example, ask everyone in the room to answer some questions (i.e., fields or variables). What is your gender? How old are you? What color is your hair? How much do you agree with items like, I love cars, I love dogs, I love Chinese food? Your answers will be unique to you—your descriptor. Furthermore, some of your answers will be the same as others in the

room and different from some others in the room. Put them all together, and you have a descriptor set that you can use to group and compare subgroups based on how people responded to the questions.

> **BRUCE'S TIP #23**
>
> Using Descriptors
>
> You may not always use descriptors, but it is good to think about whether or not they might strengthen your analysis before you begin.

As a powerful tool, descriptors allow you to break out your qualitative work across demographic and other survey type data for greater insights. When using descriptors in your qualitative data analysis, in addition to being able to analyze media files, you are able to break out the information that describes the sources of your media files to see things from different perspectives and introduce new dimensions to your analysis. Since Dedoose allows for multiple sets of descriptors, you can add as many levels of analysis as your study needs.

As discussed earlier, descriptors are the characteristics of the participants in your research (e.g., individuals, dyads, families) but can also be descriptions of settings in which observations are made (e.g., stores, schools, neighborhoods, cultures). In Dedoose, "Descriptors" are sets of information you use to identify and describe the sources of your media (e.g., documents, video, audio, images).

Descriptors are information that describes the source of your data (e.g., research participants, families, schools, other settings) at a particular level of analysis. The descriptor fields or variables (the term *fields* is used in Dedoose) that constitute each descriptor set may include demographic information, census data, dates, scores from survey measures, test results, and any other information you gather that is useful in describing and distinguishing the source of your media/data—essentially your level(s) of analysis.

- *Topical scope:* statistical study (breadth, population inferences, quantitative, generalizable findings), case study (depth, detail, qualitative, multiple sources of information)

- *Research environment:* field conditions, lab conditions, simulations

- *Descriptive studies:* Who? What? When? Where? How much? Descriptions of population characteristics allow for estimates of characteristic frequency and the discovery of associations among variables.

Figure 6.1   Multiple Descriptor Sets

Figure 6.1 shows a study with multiple descriptor sets, one for each different level of the data. Here, some of these descriptor sets can be consolidated into a single set. However, if your research questions only focus on data at the participant level, then the study is more complicated than it needs to be.

For instructions about how to create descriptors, see Section 8.1.1 of Chapter 8.

6.2.1 Multiple Descriptors

The use of multiple descriptor sets is ideal for studies that address question(s) that look across different levels of analysis. For example, if a study is comparing student outcomes across different school districts, there might be three levels of data: (1) the district, (2) the schools within each district, and (3) the students within each school. Commonly, different sets of descriptor fields/variables would be collected to distinguish the cases at each level (i.e., district-level fields might

include average family annual income, square miles of capture area, percentage of rural vs. urban neighborhoods; school-level fields might include size of student population, student–teacher ratio, percentage of children on free lunch program; and student-level fields might include age, gender, grade level, size of family, language spoken at home, and standardized test scores).

Furthermore, imagine that the study's data come from interviews with parents about the home educational environment. Under these circumstances, each interview could be linked to three descriptor fields or sets: (1) the one specific to their child, (2) one for their child's school, and (3) one for their school district. Dedoose allows exploration of qualitative data and coding activity across multiple descriptor levels. Thus, in a study like this, variations in the qualitative data and coding activity could be explored as a function of district, school, or student fields and combinations of fields across these levels.
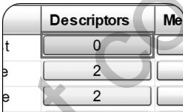
6.2.2 Dynamic Descriptors

If you are doing a longitudinal study gathering data from the same sources at multiple time points, dynamic descriptors are how you will be separating and organizing these time points. By assigning a field as dynamic, you are allowing it the flexibility of having multiple values based on what media file it is linked to.

BRUCE'S TIP #25

Dynamic Descriptors
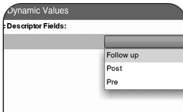
Given proper reason to use them, dynamic descriptors can be very helpful to your analysis.

Locate a media file that needs to be linked to update

Create and link, edit a linked descriptor, or link an existing descriptor

Assign a value for the dynamic descriptor field you created

To create a dynamic descriptor field, create a new field ensuring you check the "Dynamic Field" box. Once created, editing or linking a descriptor from this descriptor set in the "Descriptor Links" or "Linked Media" windows will prompt you to assign a dynamic value specific to the media.

Most descriptor data are static, in that the characteristics of research participants usually do not change over the course of a study. Dynamic fields are primarily designed to track change in your data over time in a longitudinal study. Dynamic fields are typically used sparingly, and the values for these fields are set each time a user is linking or editing a static descriptor to media (one-to-many relationship). Once properly linked and assigned to your media, these fields are able to be analyzed as any other static field would be, the main difference being that media linked to the same descriptor can fall into different categories. For example, when analyzing the preinterview option of a given dynamic descriptor field, you would only be viewing analysis sourced from media in the preinterview phase of your study, as opposed to all media linked to a specific participant's descriptor.

In this section, we have talked about strategies for mixing your research design and analysis and about descriptors. Using descriptors takes you beyond coding and allows the researcher to categorize their data in different ways. You may not always use descriptors, but it is good to think about whether or not they might strengthen your analysis before you begin.

6.3 Topic Modeling

Rapid advances in technology continue to challenge the role of the researcher in the analysis process. To support these changes, Dedoose is developing an interactive topic modeling engine prototype. This approach to analysis is driven by human interaction with the algorithmic model, both training and leading the inquiry. This new functionality will likely be embedded and integrated into the Dedoose environment in 2020. Although this function is not yet available, the following discussion about using topic modeling as your approach to mixed methods data analysis offers valuable insight into how you might develop your analysis thinking. Also, by giving thoughtful consideration to how you approach topic modeling analysis, you can better assess your overall skills in mixed methods research.

This different type of analysis develops through the research supervising iterative interactions with the topic model algorithm that produces the data analysis model. The researcher can then evaluate the model and give feedback about the analysis, which the algorithm then integrates into its processes and presents a new model for review. The researcher can keep running these iterative cycles until they are satisfied with the results. See Section 6.3.4 for more details on this interactive process.

6.3.1 Background

Qualitative and mixed methods researchers are increasingly using data apps with the intention of improving the quality and efficiency of analysis. Early attempts by developers to respond to this growing interest has introduced a range of technology tools. Examples include word frequency count, auto-code by theme, auto-code by sentiment, and automatic grouping of similar sentences for cluster extraction.

Caution is warranted, however, when using any automated app function such as these to analyze data en masse. Despite its attractiveness and convenience, automation may distance or remove the researcher from the analytic process. One consequence of this distancing or removal is the limitations placed on the researcher's ability to monitor or evaluate results. This critical limitation would possibly allow the introduction of various coding errors being introduced into the project data. The challenges and pitfalls of these tools are therefore of serious methodological concern.

Discussed here are some concerns about what technology can provide from the perspective of sound qualitative and mixed methods design, implementation, and rigor. Despite their possible shortcomings, there are many reasons a researcher may desire a technological assist. Efficiency in managing large amounts of data and calls for transparency and quality in analysis are at the forefront of this trend. As a result, efforts have been made by some developers to provide this functionality. For example, qualitative data products such as NVivo, QDA Miner, Max QDA, and Atlas.ti offer features for sentiment analysis, topic detection, language detection, keyword extraction, and entity extraction. Web-based services such as MonkeyLearn also offer text analysis with machine learning to turn tweets, emails, documents, webpages, and more into actionable data. They promote a business model for automated processes intended to save hours of manual data processing.

How well these features work depends on the nature of data involved, which can include sets of open-ended narrative interviews, responses to more structured survey questions, social media streams, news and other articles, and archival documents. What do these features do well and what limitations exist? How does such activity affect the quality or depth of one's analysis? What may be coming next?

Software developers are continuously exploring new ways in which digital tools can immerse the qualitative researcher in a deeper, more personal, connection with the data. Technology offers pathways of connecting a wide range of senses with the analysis process. Imagine hearing the voices and seeing the setting from multiple views while you code the text.

Developing an informative and comprehensive code system for a project is central to producing rigorous and valuable research or evaluation project findings. Under most circumstances, and certainly historically, this was a person-based process and there are a variety of approaches to identifying themes in qualitative data that will best serve a project's goals (Ryan & Bernard, 2003). However, with the advent of technologies to support this work, developers have searched for ways to automate this process as much as possible, removing the manual burden of reading text, identifying themes/codes, determining meaningful segments, and then applying codes. The first efforts to address this challenge focused on the ability of computer systems and programs to identify keywords, entities, and sentiment in natural language. In general, this coding process is one of classification. That is, researchers try to group what is found as meaningful information contained within qualitative data into a pattern or theme that helps in understanding the multiple meanings represented in the data. Attempts to automate this process rely on the capabilities of computer-based classification systems, often referred to as text mining, machine learning, and natural language processing. Much of this work relies on the identification of word and

word phrase frequencies and probabilities. While this work shows great promise, it is generally understood that natural language is complicated when considering how words are used in different contexts, for example, sarcasm, local vernacular, and the use of metaphor introducing complications that can be lexical, syntactic, semantic, and pragmatic (Anjali & Babu Anto, 2014). In describing the relative effectiveness of existing tools, "looking at these products it is clear that any sort of truly automated processing of text is in the future, except in some highly restricted domains where controlled vocabularies can be used" (Blank, 2008, p. 547).

6.3.2 Modes of Inquiry

As an illustration of the utility of these tools, this section presents the results of a pilot project using MonkeyLearn as a tool to critically explore features for keyword, key phrase, and entity extraction, sentiment analysis, and subsequent auto-coding. A review of the current state of these automated functions frames the discussion with particular attention on the use of these features, what they can offer the end user, and where they might be misused.

For this pilot project, MonkeyLearn (https://monkeylearn.com), a web-based services for text mining and automatic categorization, was integrated into the Dedoose data management and analysis environment. A total of 17,389 *American Journal of Public Health* abstracts were processed and auto-coded in Dedoose. The results included 2,178 codes, which represented 240 entities (137 locations, 10 people, and 93 organizations), 1,934 keywords (or key phrases), and 4 sentiments. These "codes" were automatically applied 72,725 times to 27,896 independent excerpts. Without question, this was a tremendous processing feat that was completed within a matter of minutes. However, from the perspective of what represents high-quality and rigorous qualitative research methods, the results were far from satisfactory and of little immediate usefulness.

A preliminary examination of the results revealed, for example, that among the 1,934 keyword or key phrase codes identified, each needed to be found three or more times in the total corpus to emerge. Thus, auto-coding provided a reasonable job of word, entity, person, and sentiment classification (Namey, Guest, Thairu, & Johnson, 2008) based on the presence, or absence, of words and word phrases. However, results from the pilot study show that an efficient process is needed to reduce the overall number of codes into a useful and manageable system.

The pilot project also examined meanings based on context and related thematic decisions about code definition and application. Results identified 27,896 excerpts with one or more codes that required further analysis. Coding based on only the presence of words or word phrases is likely to classify an overwhelming number of excerpts into a system where a human would not be fully satisfied with the results from a contextual perspective. Thus, to thoroughly "clean" the data set, each excerpt must be reviewed, and any inappropriate codes removed.

That said, there are certainly many controls, like stop word and synonym lists, and minimal criteria for defining codes based on words or word phrases that could have been put into place prior to data processing that would have significantly

improved the initial automated results. Nonetheless, from a purely qualitative methodological lens and consideration of the complexity of natural language, the research community must take care when conclusions are drawn from findings where auto-coding is based on keywords, entity, and sentiment extraction alone.

6.3.3 Recent Developments

Fortunately, there is promise in more recent developments within the computer science community around topic modeling in qualitative research. Topic modeling is a rapidly developing area in machine learning, text mining, and natural language processing. Particularly at a time when researchers are looking to harness the value of "big data," there is a great deal of exploration into how these models can be used in many areas of inquiry, including the social sciences. Topic modeling aims to automatically identify "topics"—that is, organizational (semantic) structures within a corpus of documents (or set of excerpts from documents)—that reveal key themes around which the corpus can be understood (e.g., Latent Dirichlet Allocation; Blei, Ng, & Jordan, 2003). From a technical perspective, the process relies on the probabilities that words co-occur within individual documents in the overall corpus. For example, in a set of summaries about animals, the words "feline," "fur," "lick," and "purr" were often found in the same summaries and the words "scale," "swim," "water," and "fin" often occurred together. A topic model might identify two distinct classes, of which a person might identify one class as including "cats" and "fish." This may seem oversimplistic, but algorithms for learning topic models are designed to exploit these word co-occurrence structures and find the topics that best describe and organize the corpus of documents. The resulting topics are distributions over the entire vocabulary of words but can also be used to obtain a representation of each document as a distribution over topics. Essentially, a text mining approach is to identify, on a probability basis, how likely it is for sets of words to occur within individual documents. Furthermore, while individual documents will often fall into multiple topics, they will do so with greater or lesser probability. Fundamentally, like qualitative content or thematic coding, topic modeling is a data reduction strategy. It seeks to understand and reduce the number of dimensions or themes across an underlying or latent organizational structure categorizing the important meanings expressed through the natural language in documents.

The application of topic modeling in traditional qualitative and mixed methods research is both promising and challenging. The promise includes the possibility of far more rapid organization of "big data" where unstructured text is in play and, perhaps, the ability for an investigator to argue that a less subjectively defined conceptual framework has been discovered given a purely word-based derivation of the classification system. This "out-of-the-box" application of topic modeling is far more sophisticated than auto-coding based on word and entity frequencies or sentiment ratings as can be found in other contemporary software. Furthermore, as these models are mathematically derived and essentially arbitrary with respect to the real "meanings" conveyed in natural language, it is often the case that topics found by the algorithm don't satisfy the needs of the researcher performing the analysis. Another factor is the nature of the

data involved. Again, like many mathematical models, the more data there are to build the model, the better the likely outcome. So these algorithms perform better where the corpus of data is enormous. However, in many areas of social science research, data sets are often relatively small. Thus, the challenge of "thinner" data in many academic research projects presents a unique challenge that must be addressed if there will be any confidence in the results of a study using topic modeling.

As work to build a useful topic modeling feature in Dedoose has progressed, both larger and smaller data set circumstances have been kept in mind and guided the design of a modeling engine and interface that will serve a wide range of circumstances. More recently, there have been various developments in interactive topic modeling, an area that explores ways with which the user can shape the resulting topics according to their needs in interactive ways. These developments include the idea of machine learning, supervised learning, and iterative topic modeling. If communities of qualitative researchers consider embracing the use of statistical algorithms in their work, there must be an adequate level of transparency in the procedures and, at best, clear levels of human control. More recent developments into working and refining topics such as "tuning," "splitting," and "removing" topics may offer the researcher this transparency and control, thus instilling the trust that researchers strive to attain to have confidence in the methodological rigor in our work (Gibbs et al., 2002; Patton, 2015; Poulis & Dasgupta, 2017; Salmona & Kaczynski, 2016; St. John & Johnson, 2000).

In short, given the insights that can be provided after running a topic model, researchers, with their unique human ability to interpret meaning in context and avoid the inherent limitations of simple keyword-based or other classification approach (Cambria & White, 2014), can train and tune the model by providing feedback in a variety of forms. The model can then learn from this feedback and be rerun and retuned iteratively until an acceptable, coherent, and interpretable structure has been defined and applied. From that point forward, the researcher can articulate the structure of the model and have confidence in subsequent classification of new data as determined by the algorithm. Topic model processing can provide a variety of information to allow for evaluation and inform areas for tuning and refining the model. For example, a model can be presented that shows top words and documents (excerpts) for each topic as well as the probability a document is well-represented by each topic. These results that both organize the documents and provide insights into the organizational structure of the model are ideal for a researcher to interact with, evaluate, and enhance the model. As the model further tuned and trained, the model itself is taught and can be run again iteratively through an evaluation, feedback, and retraining cycle until it meets the needs of the researcher.

6.3.4 Training the Model

From a qualitative data analysis perspective, Namey et al. (2008) remind us of two general approaches to data reduction: content and thematic. Content analysis is focused primarily on the presence and frequency of words and phrases toward understanding meaning in text data. Toward this end, computers have been found

very useful in generating word/phrase counts, and appropriate software can be capitalized on to automate the tagging of content based on these counts. However, given the complexity of natural language, this approach has been criticized because it does not take context into consideration. In addition, a great deal of error can be introduced when using automated tagging software features.

In contrast, thematic coding involves investigator development of sets of codes based on implicit and explicit themes discovered and applied to content in the context of the source text. Ideally, both approaches can be used in any analysis exercise. This is often the case as researchers iteratively move from the raw text to the development of more latent ideas/interpretations and back again as a full code system evolves and stabilizes before a full application of the system to the data set takes place.

The interactive topic modeling engine prototype currently being developed for Dedoose includes features for the human "tuning/teaching" as described above. Such capability addresses many of the shortcomings and concerns described and, at the same time, capitalizes both on the power of what the software can offer alongside human input into determining how the model behaves and what it can ultimately produce.

Consider what interactive and iterative topic modeling might look like (see Figure 6.2) and how a topic modeling feature in Dedoose would function. Ultimately, the goal of topic modeling in any form is to serve as an efficient data reduction activity. During Stage 1, the researcher interrogates, interacts with, and trains the model. Once the researcher is happy with the model, it can be run during Stage 2 on the entire data set, where topics and meanings can be extracted.

From a practical standpoint, the first step to using the Dedoose topic modeling engine would be to prepare your text. If using short answers from a survey, your

Figure 6.2   Topic Modeling: Training the Model

Stage 1: Researcher trains the model   Stage 2: Looking for meanings and developing findings
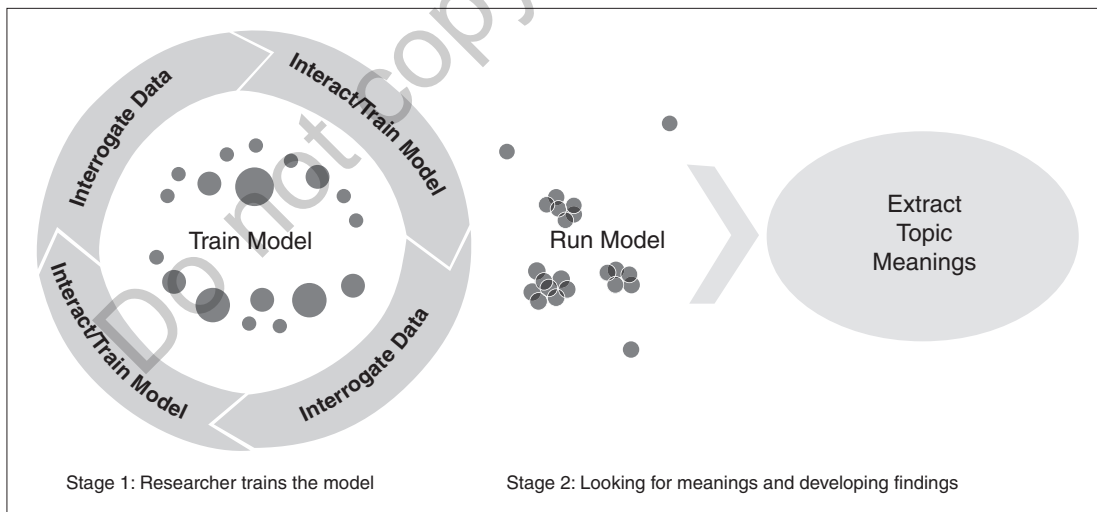
data are already prepared and can be imported directly to the system. If not, some criteria for parsing running text needs to be defined and then applied to all documents to create the excerpts that will be processed by the model. Once your data are ready and imported, you'll work through the following steps to generate and train your model:

1. Select an initial number of topics to be generated.

2. Explore top words, salient words, and top documents for each topic, and then take steps to train the model. This training can include the following:

   a. Removing words and/or documents (excerpts) and rerunning the model with those items excluded

   b. Splitting topics, which allows you to create some number of subtopics through a manual assignment of documents (excerpts) from the topic being split onto the subtopics, and then rerunning the model accordingly

   c. Reviewing removed words/documents and determining if they can be returned to an active role

   d. Merging topics

   e. Manually moving documents from one topic to another without the need to split

3. Run the final version of the trained model, export your results (at least until the engine is fully integrated into Dedoose), name your topics, and have all excerpts automatically coded based on the defined criteria.

Using topic modeling in data applications is still in an early stage. Look at the Dedoose website to find updates about how this has been integrated into Dedoose.

6.4 Case Study: Integrating Mixed Data in a Longitudinal Study

THINK ABOUT, AND ANSWER, THESE QUESTIONS AS YOU READ THE CASE STUDY

1. How did this research team use descriptors to help manage longitudinal, multisource data sets? What are some ideas that you might use in your study?

2. How did the research team members integrate their data? What are some strategies for integrating (managing and analyzing) qualitative and quantitative data across time points that you might use in your study?

(*Continued*)

3. In a database that has multiple sources of data over time (i.e., different types of data, different reporters), what are some strategies you can use to manage the data to facilitate analysis across data sources?

Integrating Qualitative and Quantitative Data in a Longitudinal Study of Youth–Adult Relationships: A Practical Example Using Dedoose

Nancy L. Deutsch, Haley E. Johnson, & Mark Vincent B. Yu

In this case study, we begin with an overview of the Youth–Adult Relationships (YAR) study and describe how we set up a database in Dedoose to manage the data. We discuss the ways in which Dedoose was used to facilitate qualitative and mixed data analysis, reflect on the strengths and challenges of the application for our purposes, and consider what we may have done differently in retrospect. We focus on some of the challenges of integrating our data, such as their longitudinal, multireporter nature; the volume of data, and number of researchers and analytic projects; choices about how to use quantitative data in relation to the qualitative data; and the use of codes for both organizational and interpretive purposes.

YAR is a longitudinal, mixed methods study of adolescents' relationships with important nonparental adults. Spanning more than 3 years, a variety of quantitative and qualitative methods, and multiple researchers, the study seeks to understand the role of supportive youth–adult relationships (aka VIPs) in adolescent development. We focus on five areas of inquiry: (1) understanding how youth develop and sustain relationships with nonparental adults, (2) the characteristics of those relationships and adults whom youth identify as important, (3) how relationships with different adults provide different types of social support, (4) whether and how youth's relationships with adults in one setting influence youth's interactions in other contexts, and (5) whether and how relationships with important adults are associated with youth outcomes over time.

6.4.1 Setting Up the Project

The YAR research team is multilevel and over time has included one faculty and two postdoctoral investigators, three project managers, and multiple doctoral, masters, and undergraduate research assistants (RAs). Although we are located in a single geographic space, we bring different lenses to the project and work from different computers, making the teamwork features of Dedoose critical. Data collection occurred across more than 3 years, analysis is ongoing, and researchers cycled in and out of the project, making documentation and training procedures essential. Furthermore, YAR has multiple data sources (interviews, surveys, social network maps), types of participants (youth, parents, VIPs), and

time points (five total). The data set includes a total of 311 interview transcripts from youth, parents, and VIPs as well as quantitative data from surveys taken before each interview (data are described further below). Dedoose was used as the primary tool for organizing and analyzing our interview data, which was uploaded as transcripts in text form, and for mixed methods analysis.

After interviews were transcribed, the transcripts were prepared for upload into Dedoose. To reduce the need for editing transcripts individually in Dedoose, we used a three-step process to format transcripts prior to upload:

1. First, we checked transcripts for accuracy.

2. Second, we de-identified transcripts. Dedoose's platform is HIPAA (Health Insurance Portability and Accountability Act of 1996) compliant and meets the security standards for data storage required by institutional review boards, but we replaced names with pseudonyms both for ease of analysis and for an extra layer of protection.

3. Third, we ensured that there were no extraneous spaces between sentences. This avoids difficult to read output when exporting data from Dedoose into Microsoft Word and Excel.

A consistent naming process was used for all transcripts to organize the data: "Participant pseudonym-Time point-Participant Code#-Date." For parents and VIPs, a "P" or a "V" was included after the Participant Code number to denote that it was a parent or VIP interview (e.g., Frau-T1-7678464-V-7.1.14). A project manager was responsible for cleaning and uploading all transcripts to ensure consistency.

We used one Dedoose project to house both the overall study and individual researchers' analyses. Our coding and descriptor structures differentiate project-wide data and categories from those generated for individual researchers' projects (see Figure C6.4-1). When a user adds codes or descriptors to Dedoose for an individual project, the first level (i.e., "parent") code or descriptor category is labeled with the researcher's name or project title (e.g., "Yu Dissertation").

Researchers use Dedoose's "Database" feature to select only the codes or media that they need for their project or analytic task. Users can access any analysis done to date and still screen out unneeded information. This is key for a longitudinal project with multiple time points, sources of data, researchers, and topics of inquiry.

6.4.2 Data Sources

The YAR study had two phases of data collection (see Figure C6.4-2). Phase 1 was a survey of 289 youth from which we selected a subsample of 41 youth (20 middle schoolers and 21 high schoolers). In Phase 2, we followed those 41 youth for 3 years. Here, we focus on Phase 2, the phase that included longitudinal, mixed methods data, and for which we utilized Dedoose.
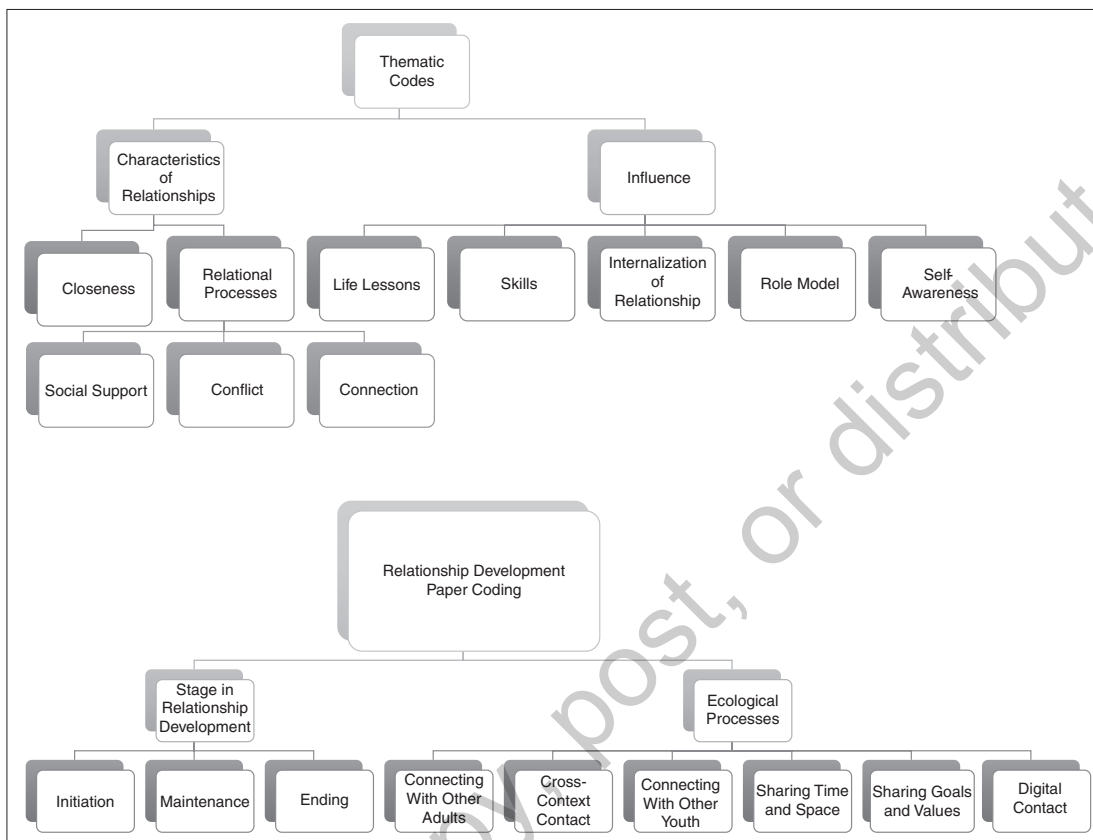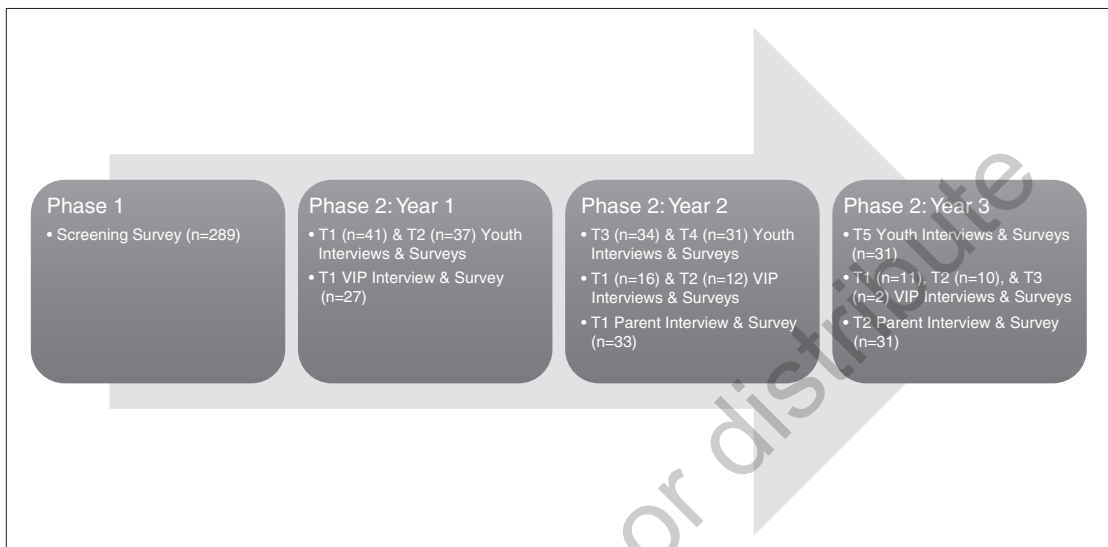
Note: Hierarchical nature of code system as shown. YAR = youth–adult relationships.

We surveyed and interviewed the 41 youth five times over the 3 years. Prein-terview surveys were administered to participants online through Qualtrics. They included basic demographic information, presence of and closeness to VIPs, and a variety of scales, including measures of Positive Youth Development (PYD; Geldhof et al., 2014), social support (Vaux, 1988), relational styles (Experiences in Close Relationship Scale [ECR]; Wei, Russell, Mallinckrodt, & Vogel, 2007), self-esteem (Shoemaker, 1980), and protective factors (Phillips & Springer, 1992). Semistruc-tured interviews included questions about the youth and their relationships with adults. In each interview, we provided a definition of "significant adult" (aka VIP) and asked if the youth had someone in their life who fit that description. If so, the youth was asked a series of questions about that relationship. If youth named a different VIP at a subsequent interview, we also asked about their previous VIP and

Figure C6.4-2   Diagram of the YAR Study

| Phase 1	Phase 2: Year 1	Phase 2: Year 2	Phase 2: Year 3
• Screening Survey (n=289)	• T1 (n=41) & T2 (n=37) Youth Interviews & Surveys • T1 VIP Interview & Survey (n=27)	• T3 (n=34) & T4 (n=31) Youth Interviews & Surveys • T1 (n=16) & T2 (n=12) VIP Interviews & Surveys • T1 Parent Interview & Survey (n=33)	• T5 Youth Interviews & Surveys (n=31) • T1 (n=11), T2 (n=10), & T3 (n=2) VIP Interviews & Surveys • T2 Parent Interview & Survey (n=31)

Note: YAR = youth–adult relationships.

why that relationship was no longer significant. At the last time points, we asked youth about their identities and goals and to reflect on all the VIP relationships they had discussed during the study.

During Interviews 1, 3, and 5, youth made social network maps, depicting different contexts of their lives and the peers and adults they interact with in those settings (see Figure C6.4-3; Hirsch, Deutsch, & DuBois, 2011). The maps included both quantitative and qualitative data; youth rated how close they felt to each person and responded to open-ended questions about the relationships on their map. Once during the study, youth made pie charts (illustrating the percentage of their interactions with adults that they would describe as positive, negative, and neutral) and graphs (plotting their relationships with their parents, best friend, and VIP in terms of levels of trust, closeness, and importance over time). The network maps were converted to quantitative data and entered in our master SPSS database. These data may be uploaded to Dedoose as descriptors for mixed methods analysis in the future. The graphs, maps, and pie charts may also be uploaded as images into Dedoose for coding as research questions drawing on those data emerge. Of particular interest will be the ability to link the sections of interviews where youth are describing their maps, graphs, and pie charts with the images themselves.

The social network map is broken up into six "slices" representing different contexts in which youth may interact with adults (e.g., family, school, neighborhood). Each youth writes their name in the middle of the map. They were

Chapter 6 | Designing Mixed Methods Analysis **151**

Figure C6.4-3 Social Network Map

Figure C6.4-3 Social Network Map
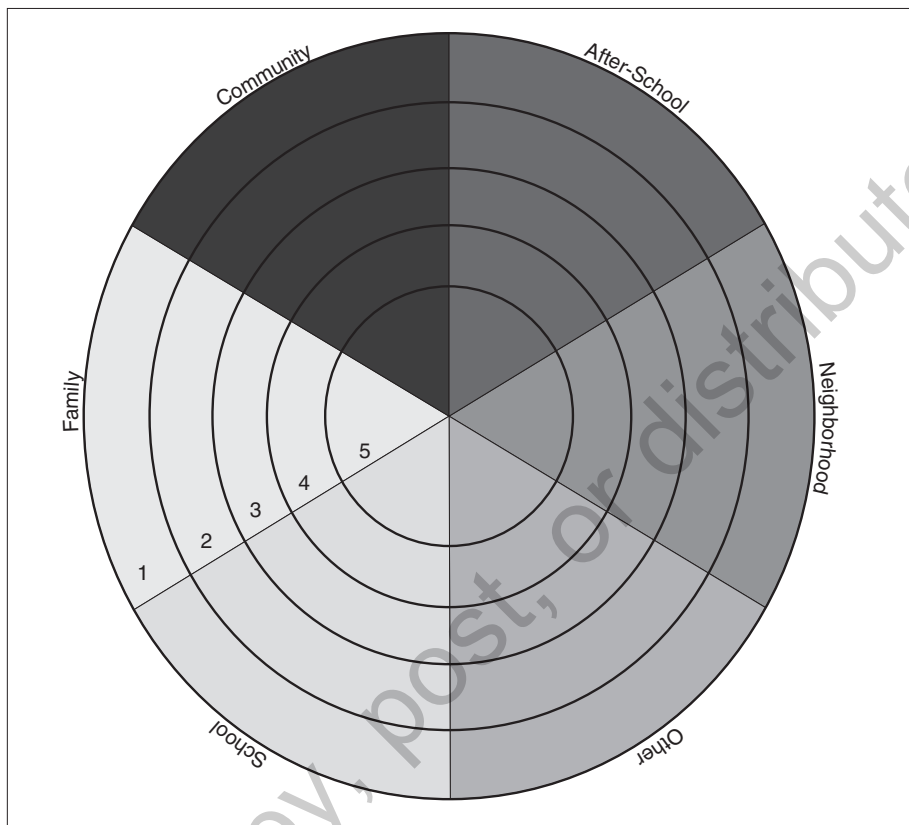
then provided with two different colored sticky notes—one for adults and one for peers. Youth were asked to place adults and peers on the map in the context in which they interact with them and at the numbered level that represented how close they felt to the person, with 5 being the closest and 1 being the least close. The names on this image are covered to protect the confidentiality of the participants and the people on their map.

We surveyed and interviewed youth's parents twice during the study and the nominated VIPs up to three times. We followed up with previous VIPs when possible. The surveys were used to assess characteristics of the adults (e.g., self-esteem, self-efficacy, relational style). The open-ended interviews focused on the adults' relationships with the study youth, with youth in general, and their own experiences with mentors.

6.4.3 Data Management

We used descriptors and codes to organize and reduce the data, making it easier to navigate the large volume of data and prepare the data for deeper levels of analysis based on specific research questions.

6.4.3.1 Descriptors: Organizing Data and Utilizing the Mixed Methods Features

Each transcript, when it was uploaded to Dedoose, was assigned a number of descriptors to categorize and describe the data and the participants (see Figure C6.4-4).

We used three organizational descriptors to facilitate sorting data in different ways: (1) Youth Anchor, (2) Wave, and (3) Participant Type. Because each youth had multiple data sources (up to five youth interviews; two parent interviews; 0–3 VIPs, each of whom was interviewed up to three times), we tagged each interview with the "Youth Anchor" descriptor to identify the youth participant with which it was associated. This allowed us to easily access all the data on a particular participant. For example, if we wanted to review all of Bob's data, we could use the "Youth Anchor > Bob" descriptor to quickly pull all his interviews, all his VIPs' interviews, and all his parents' interviews.

"Wave" referred to the period of data collection, sorting transcripts by time point. This allowed us to look for developmental patterns, something we discuss further below. It also provided quick access to all the data from a particular wave, which is important because different topics were discussed at different time points. For example, at Wave 4, we asked youth about their identity. A researcher who wanted to focus on youth identity could use the wave descriptor to quickly select the Wave 4 transcripts for analysis.

Finally, we tagged interviews by Participant Type (i.e., youth, parent, VIP). This facilitated selection of data from specific types of participants (e.g., all VIP interviews). It also enabled us to use Dedoose's "Codes × Descriptor" analysis function to compare data and codes across participant types (e.g., comparing how youth vs. VIPs talk about "closeness").

Descriptors also provide a tool for exploring interactions between our quantitative and qualitative data. To create descriptors from our quantitative data, we cleaned the survey data in SPSS and imported it into Dedoose. We found that the most effective way to use our quantitative variables in Dedoose was to create "high" (1 SD or more above the mean), "middle" (within 1 SD of the mean), and "low" (1 SD or more below the mean) groupings for each variable. We could then use Dedoose's analysis tools (e.g., the "Codes × Descriptor" chart) to compare patterns across participants more efficiently than with individual mean scores. This may not be ideal for every scale, but we used it as a starting point to examine descriptive patterns. Researchers can then create more nuanced descriptors as needed.

Figure C6.4-4 Descriptor Organizational Chart and Example Descriptors in YAR Study
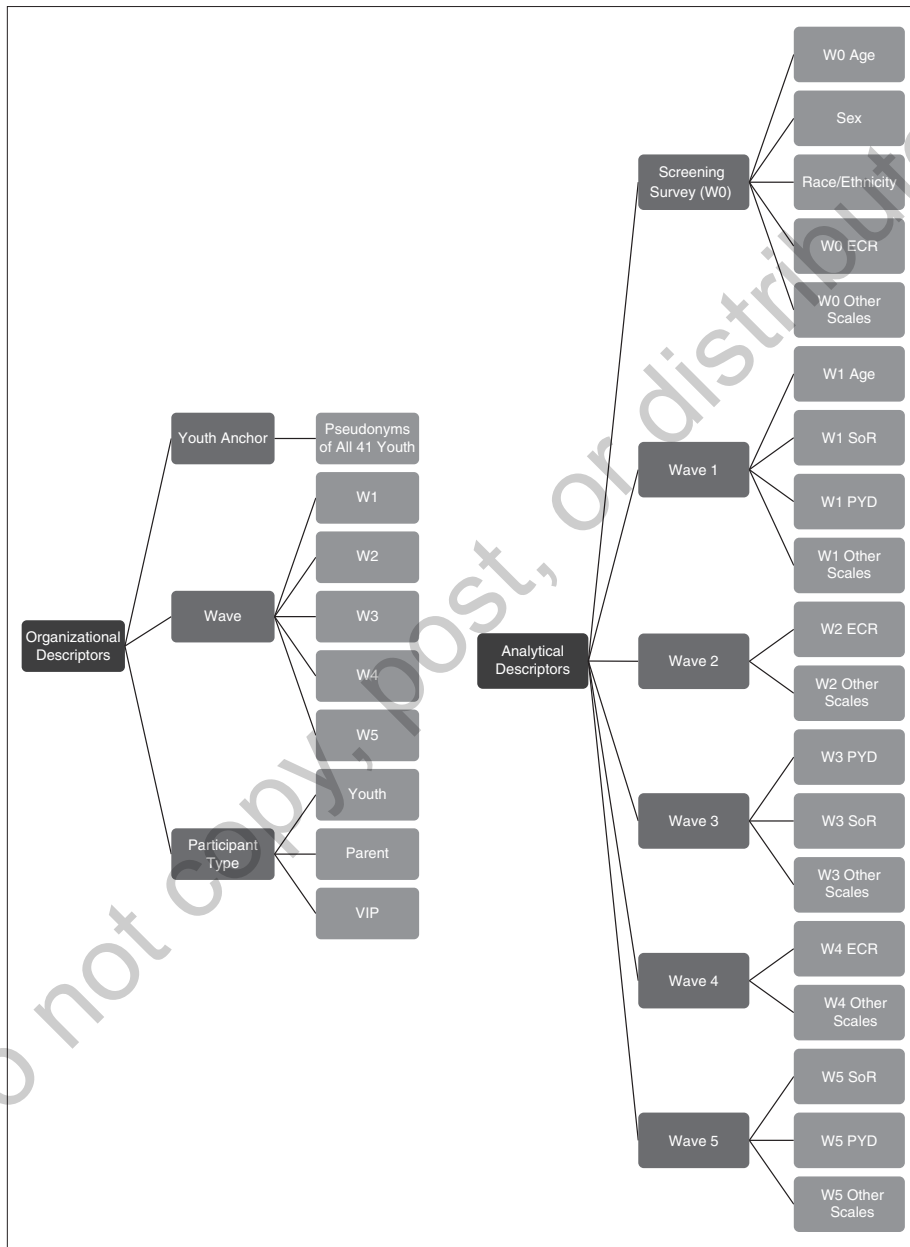
Note: YAR = youth–adult relationships.

When a youth interview transcript was uploaded to Dedoose, it was tagged with the descriptors from that youth's screening (W0) survey (e.g., ECR, PYD). These data serve as baseline youth characteristics. Each interview was also tagged with the descriptors from that Wave's survey. For example, a Youth Wave 3 interview would be tagged with both the screening (W0) survey and the Wave 3 survey descriptors. Tagging all youth interviews with baseline and time point–specific descriptors allows us to compare data based on youths' initial or current characteristics and to explore change over time.

6.4.3.2 Coding

Our coding structure includes two types of codes (Miles, Huberman, & Saldana, 2014): (1) descriptive codes, which label the topic of an excerpt based on its content (e.g., relationship with VIP), and (2) interpretive codes, which apply meaning to the excerpt in relation to the researcher's conceptual frames (e.g., social support). Initial codes were developed based on our research questions and important constructs from the literature (e.g., adult's relationship to youth; context of relationship; characteristics of the relationship such as closeness, conflict, and types of interactions; characteristics of people; influence and outcomes). We created three categories of codes to organize our data: (1) question codes (descriptive), (2) who codes (descriptive), and (3) initial thematic codes (descriptive and interpretive). We also created codes for "great quotes" (as recommended by Dedoose). These codes served as organizational buckets into which we sorted data for analysis in relationship to specific research questions. We keep track of codes in a codebook in Excel, where each category of code has its own tab with definitions and examples.

Question Coding. Coding the data by interview question is useful for data reduction and allows us to easily access all youth's responses to a given question. Each transcript was broken into excerpts using codes representing the interview time point and question number (e.g., T1-Q1). If someone is interested in understanding how youth define closeness, for example, the researcher can start by pulling all the data related to the interview question "How do you define feeling close to someone?" (T2-Q44). The question codes are applied as their own excerpts, separately from thematic coding.

Who Coding. The "who" code identifies the person about whom an interviewee is speaking. We coded for both the person's role within our study (e.g., VIP, peer, other adult) and their relationship to the youth (e.g., parent, sibling, school-based adult, nonfamilial peer). Our "who" code tree includes child codes identifying whether an adult is a current or past VIP and the time point at which the VIP was nominated (e.g., Current T1, T2 VIP, etc., see coding table). This facilitates data reduction, allowing researchers to select only the data that are relevant for their questions. For example, if a researcher wants to examine how youth are describing their VIP relationships at Time 3, they can select all the data coded as "Who → Role → Current VIP → Current T3." If a researcher wants to look at

how youth at Time 4 describe *past* VIP relationships, they can create a data set of transcripts tagged with the Wave 4 descriptor and select data tagged as "Who → Role → VIP Relationships → T1 VIP and T2 VIP and T3 VIP." This time-specific VIP coding became necessary as we added waves of data and realized that we needed a way to differentiate conversations about current and past VIPs. Like the "question" codes, "who" codes were applied as their own data excerpts to reduce the cognitive load on researchers, who applied the "who" codes in large excerpts during an initial read through of the data. This does pose challenges for searching the data for coding overlap, however, an issue we discuss below.

Initial Thematic Coding. Our thematic coding scheme was developed through a series of inductive and deductive practices (see Futch Ehrlich, Deutsch, Fox, Johnson, & Varga, 2016, for more details). Thematic codes were applied to smaller excerpts of the transcripts than the "who" or "question" codes during a second, in-depth reading. We initially planned to train RAs on the coding protocol, ensure intercoder reliability, and have coders individually code the interview transcripts. We set up a coding training in Dedoose, with the principal investigator and the coprincipal investigator as the master coders. However, because of the nature of our coding process (e.g., coding large excerpts with the "who" code before tagging smaller excerpts with thematic codes) and the structure of the Training Center (which offers up excerpts in a random order and doesn't take into account overlapping excerpts), the training infrastructure didn't work for our needs. Instead, we set up a separate project in Dedoose in which new RAs were trained and could practice thematic coding. The training database allowed RAs to become comfortable in Dedoose and with our data, without compromising the master database. RAs could familiarize themselves with our coding norms by going through the transcripts with the coding visible. They could also practice coding by turning off existing coding in the data set selector, creating and coding excerpts themselves, and then comparing them with the existing coding. After reviewing the codebook, reading through existing coded transcripts, and discussing coding at team meetings, RAs would independently code a transcript with the existing coding turned off. Codes would then be compared and discussed at a team meeting. After multiple rounds of whole-team coding training, we assigned two RAs to each interview transcript.

RAs coded transcripts independently, using Dedoose's data set feature to select only themselves as a user so they could not see the other RA's coding. After both RAs coded a transcript, they met and compared their coding, reconciled any discrepancies (see Hill et al., 2005, for discussion of consensus coding), and deleted one of their sets of codes so that each interview had one complete set of coding. Any discrepancies that they could not reconcile through discussion or which they were unsure about were brought to the full team for discussion.

6.4.3.3 Naming Protocols in Dedoose: Advice for Facilitating Data Management

Data management is facilitated by clear and consistent protocols for document and code naming. We developed the following suggestions for naming based on

our experiences with our database. First, be purposeful in assigning names to your media (i.e., documents, video, etc.) when you upload your data to Dedoose. Using a thoughtful naming protocol can make it easier to sort your media in ways that are useful to your project. For example, we named our transcripts starting with the participant name to allow for easily sorting data by participant, as the participant is the "anchor" in our study and the most common way we want to scan our data. Second, when developing code names, be sure each one is distinct. It becomes difficult to distinguish data if codes share the same name. For example, in our thematic coding, we coded for emotional support a theme that runs throughout our interviews. But we also ask an interview question about emotional support. Because of how we initially labeled the question and thematic codes, when we exported data into Excel, we had two columns labeled "emotional." This was confusing, and as a result, we renamed the codes.

6.4.4 Analysis Processes

We began analysis following the first wave of youth interviews and used ongoing reflections and discussions to shape our analysis in response to emerging questions and findings. Our team met weekly to discuss analytic tasks (e.g., coding questions and reconciliation). We had biannual data retreats in which we spent 1 to 2 days looking at data, identifying emerging themes and questions, discussing findings, and planning next steps for analysis. As researchers on our team pursued different research questions, they used the initial coding in Dedoose to select the data relevant to their question and conduct further analysis. Below, we provide three examples of what this looks like in practice.

Cross-Sectional Mixed Methods Analysis. Early on in the study, we noticed that youth were not always describing their relationships with their VIPs as "close" in the ways that we expected. We decided to use the T1 interview data to examine how youth define closeness in VIP relationships. To do this, we needed to select all the data that had been coded as both "closeness" and "T1 VIP" from the T1 youth interviews and create a data set with those excerpts. We faced a challenge in this, however, as "who" (i.e., VIP) and "thematic" (i.e., closeness) codes were applied in different excerpts (as described above). To address this, we developed a system to export data and isolate overlapping data excerpts that were coded as both closeness and T1 VIP.

First, we selected the T1 VIP and Closeness codes using the Dataset function. Then, we used the Code Co-Occurrence feature in the Analyze function, selected "include overlapping excerpts," and exported those data to Excel. We read through the excerpts in Excel and deleted any irrelevant data (e.g., large excerpts coded as T1 VIP but not closeness). We then exported relevant excerpts to Word for holistic read-throughs to identify emerging themes. This was a stylistic preference on our part. This read through could have been done in Dedoose, but the researchers conducting the inductive analysis felt more comfortable reading through data in Word or on paper to develop the codes before applying codes in Dedoose. After the themes were identified and defined, RAs returned to Dedoose and coded the

data for those themes. Finally, we used the "Code × Descriptor" function to examine patterns of difference in youth's descriptions of closeness for youth who were high, average, and low on connectedness, one of the scales from the T1 survey (see Futch Ehrlich et al., 2016).

A similar type of analysis was conducted by the third author (Yu et al., 2019), who used descriptors to sort youth into groups representing more positive and more negative attachment styles (based on their ECR scores from the screening survey). He then compared the groups on the prevalence of five types of social support (Wills & Shinar, 2000), as coded for in the interview transcripts. He found differences in the prevalence of emotional and validation support between youth with more positive and youth with more negative attachment styles. Therefore, he selected data excerpts coded for emotional and validation support and worked with a team of RAs to read through the data and identify themes, for which they coded in Dedoose. Youth with positive and negative attachment styles were again compared, this time on the prevalence of the themes identified within emotional and validation support.

Longitudinal Data Analysis With Between- and Within-Youth Comparisons. Currently, we are analyzing youth perspectives on the development, maintenance, ending, and impact of relationships across time. We began by selecting all the youth interview data, from all five time points, that have been coded as Story of Relationship. We developed an initial codebook based on the literature and our knowledge of the data and refined it through three researchers reading the data, discussing the codes, and separating and merging codes until agreement was reached on a final set of codes and definitions. We then attempted to apply the codes in Dedoose. In doing so, however, all three researchers agreed that because excerpts appeared in a random order for coding, we were losing the developmental narrative of the relationships, which was part of the question we were seeking to investigate. Therefore, we exited Dedoose, returned to the raw transcripts, and cut and paste from them to create a Word document for each youth with all their VIP data, color coded by time point. We then created a second Word document for each youth–adult relationship with separate sections for initiation, maintenance, ending, and impact, into which relevant data from the first document was copied and pasted with the color coding intact. This allowed us to (a) read holistically, for each youth–adult relationship, the story of how the relationship began, was sustained, ended, and the influence the youth felt it had for them, and (b) examine patterns of change and stability in those over the 3 years of the study. Finally, we created a third Word document for each relationship in which we applied the previously identified codes to each section of the narrative (i.e., beginning, maintenance, ending, impact), selecting excerpts of data as evidence of each code that is present and writing a brief narrative summary of the relationships. This allows us to look for thematic patterns within youth over time, within youth across relationships, and across youth within relationship stages (Arbeit, Johnson, Grabowska, Mauer, & Deutsch, in press). This analysis could have been done in Dedoose by coding for the relationships stages and then for the emergent themes. We preferred, however, to read the data holistically in Word rather than as excerpts.

Using Dedoose in combination with Word in this example facilitated the process of data reduction (Miles et al., 2014) in conjunction with a more narrative approach to analysis (Josselson, 2011). Because the data were tagged with time points in Dedoose, we were able to retain that tagging in Word to ensure that a developmental lens was applied to the analysis. As we continue, we will be able to use the mixed methods charts in Dedoose to explore patterns within and across youth over time by examining the prevalence and content of codes in relation to scores on the quantitative measures.

6.4.5 Looking Back

Dedoose was critical to our ability to organize and manage a large amount of multimethod, multisource, longitudinal data. Yet we also identified places where Dedoose was limited given our aims. For example, we would like to be able to create data sets within Dedoose based on descriptors in the same manner that you can create data sets by other organizational categories (e.g., users or codes). There are also tasks that we would approach differently now that we have been working with the data. For example, we split our organizational and thematic codes into different excerpts, as described earlier. Yet because the architecture of Dedoose relies on excerpts as building blocks for its analytic functions, searching data for coding overlap is most efficient if all codes are applied within a single excerpt. Although you can tell Dedoose to include overlapping excerpts in its searches, this approach was not optimal for our project. If we had it to do over again, we would probably put all codes into single excerpts. In addition, we found that when we wanted to look at the data more holistically or use a narrative approach to analysis, the best approach for us was to utilize Dedoose's organizational functions to identify the data we needed and then to export the data. This allowed us to read it in blocks of shared narrative meaning or structure instead of small excerpts or entire transcripts. Sometimes, we returned to Dedoose for additional coding and analysis after that step, and at other times, we did not. Thus, sometimes we used Dedoose as the main analysis software, conducting both qualitative and mixed methods analysis within the program. Other times, we drew on it primarily as a data organization and reduction tool to allow for analysis outside of the program. Yet in all cases, the program allowed us to manage and access a large amount of longitudinal data in different ways. Our lab motto is "encourage the platypus," a reference to looking for the unexpected and maintaining a creative and open mind-set when analyzing mixed methods data. Dedoose facilitated this mind-set by allowing us to look at our data in different ways and to explore a variety of combinations of qualitative and quantitative patterns in the data.

6.4.6 Information About the Case Study Authors

Nancy L. Deutsch, PhD, is a professor at the University of Virginia's Curry School of Education and director of Youth-Nex: the UVA Center to Promote Effective Youth Development. Her work focuses on how settings can best meet adolescents' developmental needs, with a focus on out-of-school time programs and relationships with mentors and other nonparental adults. Her expertise includes the use of rigorous qualitative and mixed methods in developmental research, an emphasis that she brings to bear in her work as editor of the *Journal of Adolescent Research*.

Haley E. Johnson, MEd, is a doctoral student in the applied developmental science program at the University of Virginia's Curry School of Education and a former research specialist at Youth-Nex. Her research focuses on how youth's relationships with adults can create safe spaces for identity exploration, expression of authentic self, and education in ways that promote social change and dismantling of structural inequalities.

Mark Vincent B. Yu, PhD, is currently a National Science Foundation SBE Postdoctoral Research Fellow at the University of California, Irvine School of Education. Utilizing qualitative and mixed methods, his research interests focus on socioecological and strengths-based approaches to youth development. He is particularly interested in understanding how ecological assets such as supportive nonparental youth–adult relationships (e.g., with teachers, mentors, extended family members) and settings (e.g., schools, afterschool programs, communities) can be optimized and designed to capitalize on youth's strengths.

6.5   Conclusion

Chapter 6 addressed important questions about the challenges of mixed methods enquiry and identifying appropriate analysis strategies. There was a lengthy discussion about descriptors and how they can help you when working with your data. Remember that it is not a requirement to use descriptors; only use them if it makes sense to you. The subject of topic modeling was also covered, as an introduction to the field. Check out the Dedoose *User Guide* to see more about topic modeling in Dedoose. The case study then described an ongoing project as a practical example using Dedoose. Now Chapter 7 continues this theme and delves deeper into complexity in mixed methods.