Appendix III

Collection and Analysis of Interview Data – From Qualitative to Quantitative

<u>"Qualitative" Data</u> – The data collection and analysis methods of the current study are grounded in the field of qualitative research, a mainstay of social science research (c.f. Seidman, 1998; Merriam, 2001; Strauss and Corbin, 1990). 'Qualitative' can cover any form of data that is not represented by an objective number such as a test score, grade, etc. Qualitative data should not be confused with anecdotal data, however, as qualitative data are gathered with the goal of looking for patterns in prose that are revealed through their frequency. Thus, to be useful, qualitative data should be gathered in such a way that frequency analysis becomes possible. For example, in the present study, beyond just noting that some students make comments about some aspect of their experiences in co-mentored research, the goal was to determine the frequency of those responses to determine if it is a common or infrequent theme. Student experiences in any educational situation are highly variable so it is unlikely that any theme will appear for all students, but being able to count frequency for descriptive statistics is essential.

<u>Surveys vs. Interviews</u> – Most people who seek to understand how a program is working or student experiences, turn to survey instruments to gather data. Surveys have the advantage of being anonymous, easy to distribute, able to sample large numbers of individuals, and especially useful for gathering scaled responses for quantification. They are especially useful if one is looking for an answer to a discreet question or testing well established hypotheses. At the same time, they suffer from several limitations, such as: low response rates; response biases with respect to who chooses to complete and return them; choices limited by what the survey preparer believes are what the respondents would say (but can miss the respondents' realities). When open-ended questions are provided, unless the respondent has a high level of commitment to the surveyor, responses tend to be short or missing. For text responses it also can be difficult to interpret what is meant by a response, or, even more importantly, underlying meaning or important subtexts. There is no opportunity to probe the respondent's thinking with surveys.

Consequently, interviews can provide a very rich and revealing alternative to surveys. They are, however, much more labor intensive and require great care in both gathering and interpreting data. In reality, the world is built around a constant flow of verbal data which we all interpret and base decisions on minute by minute. One has to continuously judge if information provided is truthful, a question is interpreted accurately, or an individual's response is likely to be representative for a group. From that perspective, we are all 'interviewing' others regularly. Evaluation or research interviews are systematized to allow collection and analysis of verbal data across individuals. The first key to obtaining valid data from interviews or structured conversations is for the interviewer to have no stake in any particular response, only an interest in the respondent's information. The second key is to establish a rapport with the respondent so that no particular response is sought, expected, or will have any influence on anything important to them (such as a letter of recommendation or offer of a job). Under these two conditions, an interview becomes a conversation with the neutral exchange of information. The goal and required skill of the interviewer is to simply to probe the respondent's perceptions or experiences.

<u>The Interview Method</u> – Most people are familiar with interviews used to select individuals for a job, acceptance into a school or program, choosing between applicants for an award, etc. In those cases, one is trying to compare individuals against spoken or unspoken criteria. Academic interviews most often give the interviewer great latitude in what questions are asked or even determining what kind of information to seek or emphasize. The interviewer does, however, have an internal (often unspoken or recognized) set of criteria against which individuals are being compared. By contrast, many employment interviews have evolved toward a highly structured format to make sure all interviewees are considered fairly on the same basis and the same information is obtained from each.

Interviews of the type used in this study are very different. They are designed to develop a deep understanding of a respondent's perspectives, not judge them. It requires a very different approach to the interview and skill of the interviewer to recognize when he or she may need to probe for more information. An example of this type of interview could go as follows:

Interviewer: "Tell me more about having two research mentors between NIH and the university?" (The question gives no indication of any particular response being sought.)

Student: "I really like having two different perspectives on the project."

Interviewer: "Why do you like it? What does it provide?"

Student: "They each see it differently based on their backgrounds, their interests. I get the benefit of seeing it from both directions and I get double the input. It also gives me the chance to pull both perspectives together, or sometimes decide which one I think is correct or incorrect. Sometimes one of them isn't able to see it accurately because he doesn't have the necessary background."

An open-ended survey question could be used to ask the same initial question but the response usually is the simple one line. In an interview, whether open-ended or structured, a perceptive interviewer will probe more deeply how having two mentors is affecting the learning and development of the student. The goal of the interviewer is to make no judgments or in any way lead a respondent in a particular direction.

When using interviews for data gathering, one can start from very broad questions or somewhat more targeted ones. For example, we could have simply asked each student to "Tell me your story? Tell me what you have done and how things have gone for you as a graduate student in the GPP." The conversation could start anywhere and go anywhere. This is the ultimate approach to 'discovering' what is important to individuals by what they choose to reveal. It is the right starting point if someone wants to give individuals maximum freedom to reveal whatever is important and meaningful. This method can lead to very long, meandering conversations that may or may not hit upon topics of interest to the interviewer. A respondent may have lots to say on a topic of interest to the interviewer but just never think to bring it up. For this reason, most interviews start from a series of questions that the interviewer has chosen to make sure certain topics are covered with every person. This also tends to make respondents more comfortable if they do not have to try to figure out where to start. The questions provided in Appendix 1 provide examples of this approach.

Data Collection Methods – There are several different methods of data collection one can use with interviews. First, respondents can be given questions in advance or one-by-one during the interview. The former method allows the individual to reflect, consciously or sub-consciously, on the topics to be covered in advance. By doing so, the responses during the conversation are more likely to reflect deeper consideration of the issues rather than first responses. This method may also lead to shorter conversations, as the respondent will not need to spend as much time reflecting on and thinking about their responses. If questions are provided in advance, respondents also can be asked to 'jot down' their responses to bring along as was done in the current study. This method tends to get people to take the pre-interview work more seriously when something concrete is expected of them, and their response provides a first line of data with no interpretation or prompting of the interviewer. The interviewer's role is then to ask for expansion or clarification as appropriate, and follow-up on new information revealed during the conversation.

Collection, reduction and analysis of conversation data can be very complex and often deters people from attempting to gather it. The most complete method is to record interviews so nothing is lost and a complete record of the conversation is obtained. Recording is cheap and easy but the next steps are not. To be useful, the recording must be transcribed quickly, which can be very labor intensive and expensive. It is, however, the appropriate method for gathering data for prospective research studies that have the intent of discovering patterns and forming conclusions that impact systems or processes. It also is the only method that can unambiguously establish an auditable record of the research for review by others with interest in the work.

Because this was not a prospective research study but part of a formative evaluation of the GPP, a compromise was struck between completeness of data obtained and realistic analysis. By asking students to supply written comments in advance of the interview, responses with no influence from the interviewer were obtained. Also, since the questions asked were relatively targeted, it was easy for the interviewer to simply make notes of what students added to those initial comments during the conversation. As already noted, since the goal of the Self-Study was to understand the GPP and its development with the intent for improvement, not promote it or evaluate individual students, this method worked well. Students gave no indication of discomfort or inhibition during the interviews.

Awareness must always be raised in interview studies that students may be inhibited or not answer honestly if they perceive the interviewer may be able to affect their future. For this reason, the interviewer should not be someone who makes judgments or decisions with respect to the students. (In the current study, although leading the overall GPP and its activities, neither author had significant influence over the research, academic progress or future of individual students.) One method to look for evidence of such response bias is to provide a parallel, completely confidential method of gathering information on the same topics. In this study that was done by a confidential survey including the chance to provide unstructured responses, and, as noted in the paper, there were no issues reported on the survey that was not revealed by the interviews.

Data Analysis – No matter how the data are obtained, probably the greatest challenge with interview and other text-based data is analysis. There are two distinct elements of data analysis – creation of an analysis architecture and assignment of data to the structure.

Creation of the Thematic Architecture

With most studies of this type, one does not start with a known architecture with the intent to see how the data fit. In studies of the type reported here, one usually starts with the responses of one individual to a question, chooses those responses as a starting point, and goes on to see how many of those responses and/or new ones are found in the next interview. For examples, responses of the first student to: "List the greatest benefits you see for doing dissertation research at NIH." might contain 10 different items, 5 prior to the interview and 5 more that come out while discussing the first 5. Those 10 themes become the starting point for the architecture under the topic: 'Benefits to being at NIH'. The interview with the second student might reveal 6 of those 10 and 3 new ones, and so on for all of the respondents. As more individuals are interviewed, the number of new themes emerging drops off and an idea of the frequency with which any appear becomes known. This is referred to as "theme saturation". Subsequent interviews will describe the frequency of various responses but probably not reveal any new ones that are common to many individuals. If one is determining if there are differences in responses among subgroups of a population, there must be at least 8-11 individuals from any 'group' being compared to another.

Some of the categories that emerge are obvious and easily labeled whereas others may be less obvious. For example, while talking about the benefits of doing a collaborative dissertation, one student might say: "I love the independence it gives me and the chance to direct my own research questions." This response would fit in the theme "independence". A response "I am learning how to come up with my own research questions and my mentors let me to do it." might also fit in "independence" even though the student does not use the word. Many statements or interviewer notes can fit into multiple categories. The student comment above with "mentors let me do it" would also fit into the category of "mentoring" under the theme "gives student freedom to try own ideas". Appendix II provides a full listing of the entire analysis structure that was created from student responses for the study reported here.

Assignment of Data to Analysis Architecture

Once the architecture is complete the task of actually assigning all of the data to the various themes remains. Historically, social science researchers would actually physically cut out paper versions of text data and put them into appropriate categories. Fortunately, computers have made this process a whole lot easier! The particular software package we used is called NVivo® but there are a number of similar products available. Many of them have very powerful capabilities for searching text etc., but for the purposes used here they are most useful for creating the virtual equivalent of the cutting and sorting of the past. The architecture as it is being created is put into the software program, much like a typical database architecture. Each of the text documents from interviews and answers to questions are then imported to the database and coded by the name of the individual. One by one, the documents are reviewed on the computer screen side by side with the thematic architecture. When words from the document fit with the theme, they are highlighted and dragged (or some other method) to form a virtual link with that theme. The same words can be linked to multiple themes as appropriate. This step is often referred to as "Coding". Reports can then be generated by theme to instantly bring all of the words, along with the identifying characteristics of their document, into view. This data reduction step makes it possible to look at the words linked to the theme from all or a portion of the interviews. The words can then be studied to see if they really do all fit with the theme or whether there are sub-themes to consider. This method also allows for the very important step of being able to determine how many individuals gave responses consistent with any given theme.

Another important feature of data analysis software is the ability to assign an individual to groups and compare responses by group. In the current study students were grouped in several different ways – by partnership program; number of mentors in collaborative dissertations; formal partnership vs. individual agreement; gender; and U.S. versus international student. If one were doing a similar study as prospective research, care would have to be taken to assure that sufficient numbers of individuals (8-11) were obtained in any group in order to permit a meaningful comparison. Since this was an initial formative evaluation, the design did not rise to that level for many of the potential comparisons.

Additional Considerations for Collection and Analysis of Interview Data

No matter what methods are used, it is simply not possible for an interviewer or data analyzer to eliminate his or her thinking or previous experiences when talking to someone or reading their responses. Thus, as in all research, the goal is to recognize and minimize the potential for unintended biases. As already mentioned, the most important first step is for the interviewer or qualitative researcher to set and maintain an internal desire to find information or answer a question, not go looking for something they hope to see or 'prove' an hypothesis. If that is difficult to do, it is important to employ an interviewer who is distant from the interview topic, often with little or no experience related to what is being studied. The interviewer can still be trained to listen for and probe topics of interest. With data analysis and interpretation, great caution must be exercised to resist reaching conclusions quickly based on a few

interviews, especially if a study is designed to test for a particular hypothesis or expected outcome. To diminish this risk the goal should be to look as hard as possible to find examples that do <u>not</u> fit the expected outcome, examples to disprove the hypothesis. Although inferential statistics are seldom used with interview data, the goal of trying to disprove the hypothesis as the means to build confidence that it may be true is the same. Finally, for interview-based qualitative research, it is usually expected that at least two different people with two different perspectives will read, interpret and code transcripts independently. Once independent interpretation is complete, analyses are compared and discussed to achieve consensus.