CHAPTER 3

METHODOLOGY

The population and sample of this study are identified in Chapter 3. Methods of data collection, including the development and administration of a survey are discussed. Procedures used to interview a subset of the sample are described. Both descriptive and qualitative methods were combined in this research.

Population

Previous researchers focused their work on aspects of measurement-driven instruction and high-stakes testing in individual schools or in selected school divisions in various states. However, to my knowledge, Virginia teachers had not been asked to provide their opinions about these topics in a systematic manner. Through this study, I sought to fill this gap in the research literature.

Approximately 86,000 Virginia public school teachers work in 132 city and county school divisions across the state, educating some 1,134,000 students (Source: Virginia Department of Education, August, 2000). School districts span a range of wealth from relatively prosperous schools and divisions in northern Virginia and in certain urban areas to those with far fewer resources in more rural areas of the state. While I would have preferred to define all Virginia public school teachers as the population for the study, the Virginia Department of Education (VDOE) does not keep records of names and addresses of individual teachers employed throughout the Commonwealth. The process of contacting individual teachers through principals would have been prohibitively time-consuming and would have limited access to non-respondents. Therefore, the Virginia Education Association (VEA) was contacted to request permission to survey a sample of the VEA membership for the study. Agreement to participate was subsequently received from the VEA president.

The VEA is Virginia's largest professional organization for teachers with a membership of 47,133 as of April 4, 2000 (R. Shotwell, VEA, Personal Communication, April 4, 2000). Only full-time VEA teachers from the 1999-2000 membership list were a part of the population of the study. The VEA Director of Research was named as the contact person to assist me in selecting a systematic sample of teachers from the VEA membership.

Selection of Sample

Given a population of 47,133, a minimum sample of 381 was determined by consulting Krejcie and Morgan (1970). However, a sample of 464 teachers, 25 % larger than the recommended sample, was finally chosen for several reasons. First, the larger sample helped to enhance confidence that sample responses do not vary significantly from the true opinions of teachers in the VEA membership. Second, as the 1999-2000 VEA mailing list was used to survey VEA members in the fall of 2000 (a new membership year for VEA), it was likely that some

1999-2000 members had moved or had left the education profession. Over-sampling helped to ensure that the minimum sample of 381 practicing VEA members could be reached by mail. Additionally, the larger sample size helped to ensure sufficient stratification of teachers according to the independent variables chosen: teaching experience, school SES, SOL test grade (yes, no), teaching assignment, and tenure status. And, last, a larger sample helped to strengthen the generalizability of findings from the study.

A systematic sample of teachers (Fraenkel & Wallen, 1993) was selected. Given a population of 47,133 and a desired sample of 464, the algorithm 47,133) 464 produced a selection interval of 101 to identify 464 teachers. For simplicity an interval of 100 was chosen to guide the sampling process. First, a number was randomly selected from a table representing the range between one and one hundred. That number identified the first VEA member selected from the alphabetical VEA mailing list. Next, members corresponding to each interval of 100 from the point of random start were selected until 464 teachers had been chosen. The VEA agreed to provide names and addresses for the selected members.

Instrumentation and Data Collection Procedures

Construction and Testing of the Survey

In this section the research question of the study is explained along with the survey domains and content validation procedures. Procedures for formatting and administering the survey are discussed. Exhibits are presented to describe relevant components of the survey instrument.

Research Question

The research question for the study was: In the view of Virginia teachers, how do Virginia's Standards for Accrediting Public Schools in Virginia (SOA) and the Standards of Learning (SOL) tests and curriculum affect: (1) outcomes for students, (2) outcomes for instructional practices, (3) outcomes for schools, (4) outcomes for public confidence in teachers and schools, and (5) outcomes for teachers?

Survey Domains

Survey domains were derived from two specific sources, first from the Standards of Accreditation (SOA) themselves and second from a study of related research literature. I originally defined nine domains: (1) beliefs about the effects of SOA and SOL on student achievement, (2) beliefs about the effects of SOA and SOL on instruction, (3) beliefs about the effects of SOA and SOL on school management, (4) beliefs about the effects of SOA and SOL on public confidence in schools, (5) beliefs about the effects of SOA and SOL on students' life chances, (6) beliefs about the effects of SOA and SOL on public confidence in teachers, (7) beliefs about the effects of SOA and SOL on teacher autonomy, (8) beliefs about the effects of SOA and SOL on teachers' job satisfaction, and (9) beliefs about the effects of SOA and SOL on teachers' psychological health. These domains, domain definitions, and related items are in Appendix A.

Content Validation

Content validation of the survey instrument was accomplished as follows. First, an instrument containing survey items and response sections for domain identification, item-domain association strength, and statement clarity was constructed (see Appendix B). Specific domain descriptions were developed to explain what each domain purported to measure. During June of 2000 experienced educators in a principal preparation program classified each statement into domains and rated each statement for its level of association and clarity. The educators also made recommendations for rewording of statements for clarity or accuracy. Means and standard deviations were computed for the level of association of statements with domains and the clarity of survey statements. Using this information, I was able to enhance the content validity of the survey by reallocating statements to different domains or by rewriting, rewording, or deleting statements. Content validity was tested in the same manner with a group of doctoral students in education during July of 2000 with additional refinements being made. As a result of this work, I determined that respondents were consistently demonstrating confusion when allocating statements to several of the domains. Some of the nine domains were clearly measuring the same constructs. Therefore, some of the nine domains were combined and the total number of domains was reduced to five.

Domain 1 (student outcomes) and domain 5 (students' life chances) were combined into a domain entitled *outcomes for students*. Domain 2 (instruction) was retitled as *outcomes for instructional practices*. Domain 3 (school management) was retitled as *outcomes for schools*. Domain 4 (public confidence in schools) and domain 6 (public confidence in teachers) were combined into a domain entitled *outcomes for public confidence in teachers and schools*. Domain 7 (teachers' autonomy), domain 8 (teachers' job satisfaction), and domain 9 (teachers' psychological health) were combined into a domain entitled *outcomes for teachers*. Each domain became a separate dependent variable in the study. A summary of the final five survey domains, domain definitions, and related statements are in Exhibit 1. Content validity tables are in Appendix C.

Only items associated with the anticipated domain by 80% of the reviewers became candidates for inclusion in the final survey. I made every effort to include items with the highest average mean scores for association and clarity while maintaining a numerical balance of items among the domains. Twenty items were eliminated in this manner to produce the final 52 items on the survey.

EXHIBIT 1

THE FINAL FIVE DOMAINS AND SURVEY ITEMS

Domains and Descriptions

Domain 1: Effects of SOA and SOL on *outcomes for students*.

Description: This domain will assess teacher beliefs about how SOA and SOL are likely to affect students' achievement, their feelings about school, and consequences for subpopulations of students.

Domain 2: Effects of SOA and SOL on *outcomes for instructio*nal practices.

Description: This domain will assess teacher beliefs about how SOA and SOL are likely to affect instruction in the classroom itself.

==========

Domain 3: Effects of SOA and SOL on *outcomes for schools.*

Description: This domain will assess teacher beliefs about how SOA and SOL are likely to affect schools' authority to make decisions and their freedom from outside influence.

Domain 4: Effects of SOA and SOL on *outcomes for public confidence in teachers and schools*.

Description: This domain will assess teacher beliefs about how SOA and SOL are likely to affect the faith the public has in Virginia teachers and schools.

Domain 5: Effects of SOA and SOL on *outcomes for teachers.*

Description: This domain will assess teacher beliefs about how SOA and SOL are likely to affect teachers' authority to make instructional decisions, their relative happiness with their profession, and their emotional well-being.

Domains and Items from the Survey of Likely Outcomes of Virginia's Mandated Curriculum and SOL Testing Program, September 2000 (Following final content validity study)

1. Beliefs about the effects of SOA and SOL on *outcomes for students*
    3. Students will feel too much anxiety about SOL tests.
    8. Financially disadvantaged students will fail SOL tests in disproportionately higher numbers.

(exhibit continues)

11. Students with academic learning disabilities will fail SOL tests in disproportionately higher numbers.
12. Students' writing competency will improve significantly. (R)
20. More low achieving students will be retained in school.
25. Students will have more negative attitudes about attending school.
29. More low achieving students will drop out of school.
36. Minority students will fail SOL tests in disproportionately higher numbers.
37. Student scores on nationally normed tests will improve. (R)
40. Students with low SOL scores will be tracked into low-level classes.
41. Students' reading comprehension will improve significantly. (R)
49. Students' math competency will improve significantly. (R)
51. More low-achieving students will be placed in special education classes.
52. Students will experience less joy in learning.

2. Beliefs about the effects of SOA and SOL on *outcomes for instructional practices*
    2. Teaching practices will emphasize rote recall of facts.
    4. The goal of classroom testing will be to improve SOL test scores.
    13. Teaching practices will improve. (R)
    15. Teaching practices will emphasize a curriculum that is too narrow.
    21. Teaching practices will begin to resemble SOL tests.
    30. Too much teaching time will be spent preparing students for SOL tests.
    35. Less time will be devoted to topics not addressed on SOL tests.
    42. More time will be devoted to teaching critical thinking. (R)

3. Beliefs about the effects of SOA and SOL on *outcomes for schools*
    5. Schools with low SOL scores will be told how to improve.
    6. Schools' decision-making authority will be reduced.
    14. Schools will be empowered to develop their own vision for student outcomes. (R)
    22. The SOL curriculum will give schools autonomy to make their own decisions. (R)
    28. The publication of SOL test scores will encourage schools to define their own instructional problems. (R)
    34. Threats by the state to take over low-achieving schools will cause them to improve. (R)
    43. Schools with low SOL scores will have less autonomy to make their own decisions.

4. Beliefs about the effects of SOA and SOL on *public confidence in teachers and schools*.
    1. The SOL testing program will raise the status of public education in the state. (R)
    7. The public will be more confident in Virginia's public schools. (R)
    16. Schools with high SOL scores will be perceived by the public as good schools.
    17. The public will support increased taxes to significantly raise teacher salaries. (R)
    23. Schools with low SOL scores will be perceived by the public as bad schools.
    27. Opinion polls will indicate greater confidence in Virginia's teachers . (R)
    33. The public will have greater faith in teachers. (R)

(exhibit continues)

44. More citizens will express faith in the teaching profession. (R)
45. Parents will be less likely to criticize teachers' performance. (R)
48. Public support for schools will be determined by school SOL scores.

5. Beliefs about the effects of SOA and SOL on *outcomes for teachers*
9. Teachers will experience increased self-esteem. (R)
10. Teachers in SOL test grades will ask to transfer to other grades.
18. Teachers will be less happy in their jobs.
19. Teachers will worry about their students' SOL test scores.
24. More teachers will stop teaching to accept other jobs.
26. Teacher absenteeism will increase.
31. Teachers will develop more positive attitudes about teaching. (R)
32. Teachers will experience greater stress due to the SOL testing program.
38. Teachers will experience greater satisfaction with teaching as a career. (R)
39. Teachers will worry more about their job security.
46. Teachers will worry that low SOL scores will damage their reputations.
47. Teachers will have greater authority to choose curriculum topics. (R)
50. More teachers will complain about "burnout."

[1] Statements marked with (R) were *reverse scored.* Therefore, those scored as 4 were scored as 1; those scored as 3 were scored as 2; those scored as 2 were scored as 3; and, those scored as 1 were scored as 4.

Formatting the Survey

Procedures recommended by Dillman (1978) were followed closely in the development and administration of the survey to help ensure the clarity of the instrument and to help facilitate a high rate of return. The questionnaire was prepared in tri-fold format on off-white paper.[1] Its cover was designed to include the VA Tech logo. The items, derived from Virginia's Standards of Accreditation (SOA) and from research literature, were presented in random order to help ensure that each statement was considered separately. Statements were written in a grammatically consistent manner throughout the questionnaire, and demographic questions about respondents were asked at the end of the questionnaire.

Administration of the Survey

Again, following the advice of Dillman, a cover letter was developed and included as part of the survey itself. The letter explained the purpose and importance of the study, its potential benefit to Virginia teachers, and the importance of the individual's response to the success of the study. The letter was included as part of the tri-fold document itself. Each survey received a numerical code to help facilitate communication with non-respondents; however, members of the sample were assured that confidentiality would be protected. Envelopes were addressed to individuals, and the cover letter was individually signed by the researcher and his advisor. The

[1] Linda Justus is acknowledged for her work in piloting the tri-fold format in her VA Tech dissertation: Factors that Contribute to Enrollment in Band Programs, in press.

survey with cover letter is in Appendix D. A self-addressed, stamped envelope was included with each survey.

A carefully planned mailing chronology was followed to help achieve a high rate of response (Dillman, 1978). First, letters and surveys were mailed on November 4, 2000. A postcard was sent to all participants exactly one week after the mailing to thank those who had returned their questionnaires and to remind others who had not (Dillman, p.163). Exactly three weeks after the original mailing another letter and questionnaire were sent to non-respondents. The letter explained that their questionnaire had not been received. The goals and importance of the study were reiterated as were the importance of the non-respondent's insights and answers. A self-addressed, stamped return envelope was also included. Non-respondents received a third and final mailing seven weeks after the original mailing. A similar cover letter urging teachers to participate, another questionnaire, and a return envelope were sent by certified mail.

Instrument Scoring

Teachers responded to survey statements with values of 1 to 4. Each response received a numerical weight with "strongly disagree" responses weighted with a value of 1, "disagree" responses with a value of 2, "agree" responses with a value of 3, and "strongly agree" responses with a value of 4. Higher scores indicate greater teacher agreement with statements about the likely outcomes of high-stakes testing and measurement-driven instruction while lower scores indicate less teacher agreement with the statements. Higher scores indicated a more negative view of outcomes associated with SOL and SOA.

Statements (items) marked with "R" in Exhibit 1 were reverse scored. Each respondent's domain score was the average of the item scores in a domain. Each person, then, had five scaled domain scores as a numerical summary of his or her survey responses.

Reliability

It is important for survey researchers to demonstrate that variability in responses to their surveys results from differences in respondents rather than in variability caused by multiple interpretations of items or from confusion about the items. This may be accomplished by determining the correlation of items that are thought to measure a given construct. Cronbach (1951) stated, "A reliability coefficient demonstrates whether the test designer was correct in expecting a certain collection of items to yield interpretable statements about individual differences" (p. 297).

Cronbach's Alpha (*alpha*) was computed to determine the reliability coefficients and, therefore, the internal consistency of items in each of the five domains. *Alpha* is the average correlation of items for all possible splits of the items. If item scores in a domain are highly correlated, it is reasonable to believe that respondents viewed the construct associated with the domain items similarly. High levels of internal consistency enhance the validity of conclusions drawn from survey data.

Alpha coefficients for domains measuring outcomes for students, outcomes for instructional practices, outcomes for public confidence, and outcomes for teachers were all at or above *alpha* .80 and are considered to demonstrate high levels of internal consistency. The a*lpha* coefficient for the domain measuring outcomes for schools was lower and is considered to demonstrate only a low to moderate level of internal consistency (see Table 8).

Table 8

Alpha Coefficients for Scales on the Survey of Teacher Beliefs About the Outcomes of High-Stakes Testing and Measurement-Driven Instruction in Virginia's Public Schools

Scale	Number of items	Alpha
Outcomes for students	14	.85
Outcomes for instructional practices	8	.82
Outcomes for schools	7	.65
Outcomes for public confidence	10	.81
Outcomes for teachers	13	.86

Analysis of Survey Data

Data collected from surveys were disaggregated according to the following five independent variables: (1) years of teaching experience of respondents, (2) socio-economic levels of teachers' schools (derived from free and reduced price lunch percentages provided by the Virginia Department of Education), (3) SOL test grade (yes or no), (4) level of teaching assignment (K-5, 6-8, 9-12, guidance, special education, librarian, and other), and (5) teachers' tenure status.

I believed that three variables played a prominent role in determining teacher opinions about high-stakes testing and measurement-driven instruction. First, tenure status might have played an important role in that teachers in Virginia who have completed a three year probationary period are awarded *continuing contract status*, a designation essentially the same as tenure. Teachers with continuing contracts hold their positions in perpetuity given satisfactory performance unless budgetary constraints require a reduction in personnel. And even if budget cuts were to occur, teachers with non-continuing contracts would lose their jobs before teachers with continuing contract status. Continuing contract teachers, therefore, should have less concern about periodic policy changes like those inherent in mandated curricula and high-stakes testing. The second variable that might have played a prominent role in the development of teacher opinions is the socio-economic level of their school. I presented information in Chapter 2 to demonstrate that teachers and students in low SES schools often experience more negative consequences than their counterparts in high SES schools. The final variable that might have shaped teacher opinions about high-stakes testing and measurement-

driven instruction was whether or not the teacher was teaching in a grade in which SOL testing (and subsequent public reporting) will take place. Teachers in grades 3, 5, 8, and in core high school courses are directly associated with the results of testing, both in the eyes of the public and with respect to their peers. I hoped to determine how the interaction of these variables may affect the way that Virginia teachers perceived the topics of high-stakes testing and measurement-driven instruction as assessed in the questionnaire of this study. Analysis of variance (ANOVA) was used to determine the relationships among these variables. ANOVA, like the t-test, deals with differences in sample means. However, instead of asking how only two means differ, ANOVA allows the researcher to determine whether several means differ. Additionally, ANOVA not only allows one "to deal with two or more independent variables simultaneously, asking not only about the individual effects of each variable separately, but also about the interacting effects of two or more variables" (Howell, 1987, p. 276).

The independent variables for this analysis were school SES, SOL test grade status (whether teachers were responsible for administering SOL tests), and teachers' tenure status. Three-way ANOVA was employed to describe the main effects and the interaction of variables on teacher opinions as follows: (1) the effects of tenure status (tenured, non-tenured), (2) the effects of SOL test grade status (SOL test grade, non-SOL test grade), (3) the effects of school SES (high, low), (4) the interaction of SOL test grade status (SOL test grade, non-SOL test grade), and tenure status (tenured, non-tenured), (5) the interaction of school SES (high, low), and teachers' tenure status (tenured, non-tenured), (6) the interaction of school SES (high, low), and SOL test grade status (SOL test grade, non-SOL test grade), and (7) the interaction of school SES (high, low), SOL test grade status (SOL test grade, non-SOL test grade), and tenure status (tenured, non-tenured).

F ratios were computed to compare across-group variance to within-group variance. Simple and main effects were tested in cases where significant interactions occurred (Howell, 1987, p. 405).

Construction of Teacher Interview Protocol

Protocol Design

A second phase of data collection involved interviewing a subset of respondents by telephone to provide greater depth of understanding about how the independent variables of tenure status, school SES, and SOL testing responsibility affect teacher views about high-stakes testing and measurement-driven instruction. My research in the development of the theoretical model for this study indicated that SES is likely to be a major factor in determining how students achieve, how schools function, and in determining the instructional decisions that teachers make. My experiences indicate that tenure status may affect how teachers' react to mandated curriculum and assessment requirements. Fish (1988) found that teachers in high-stakes test grades are affected differently than teachers in grades where high-stakes tests are not given. I believed that selecting teachers from the range of combinations of these three independent variables would help to ensure that opinions shaped by these circumstances would have the opportunity to emerge.

Interview questions were created for three domains after survey responses were returned and scored to produce a detailed set of directions termed an *interview schedule* (Maykut & Morehouse, 1994). After review of the descriptive statistics gathered in the study, I determined that teachers appeared to be most concerned about *outcomes for instructional practices* and *outcomes for teachers* themselves. Therefore, these two domains became a focus for interview questions. And, I wanted to know what teachers felt would be the likely outcomes of high-stakes testing and measurement-driven instruction *for students*. Therefore, this domain was included for interview questions as well.

After the three domains were written, I brainstormed words or phrases to exemplify concepts inherent in each domain. Concepts were written into a variety of interview questions and were grouped under the domain that seemed appropriate for each. Questions were open-ended to encourage a free flow of opinions from respondents (Maykut & Morehouse, 1994). A formal written protocol (schedule) including all verbal directions was prepared and read to all respondents. The preliminary interview schedule (February, 2001) is in Appendix E. The final domains and descriptions for interview questions, the interview schedule (March, 2001) and the interview protocol are in Exhibit 2.

Content Validation

Content validation of the interview questions was accomplished as follows. First, an instrument containing interview questions and response sections for domain identification, question-domain association strength, and question clarity was constructed (see Appendix F). Specific variable (domain) descriptions were developed to explain the concepts inherent in each domain. The domains and descriptions are in Exhibit 2.

EXHIBIT 2

THE FINAL DOMAINS AND DESCRIPTIONS FOR INTERVIEW QUESTIONS, THE INTERVIEW SCHEDULE, AND THE INTERVIEW PROTOCOL

Domains and Descriptions for Telephone Interviews

Domain 1: Effects of SOL testing on *outcomes for students.*

Description: This domain will assess teacher beliefs about how SOA and SOL are likely to affect students' achievement, their feelings about school, and consequences for subpopulations of students.

(exhibit continues)

Domain 2: Effects of SOL testing on *outcomes for instructional practices.*

Description: This domain will assess teacher beliefs about how SOA and SOL are likely to affect instruction in the classroom itself.

Domain 3: Effects of SOL testing on *outcomes for teachers.*

Description: This domain will assess teacher beliefs about how SOA and SOL are likely to affect teachers' authority to make instructional decisions, their relative happiness with their profession, and their emotional well-being.

Final Interview Schedule by Domain Following Final Content Validity Study (March, 2001)

Beliefs about the effects of SOL testing on *outcomes for students.*

1. What effects, if any, will SOL testing have on how students feel about school?
12. What effects, if any, will SOL testing have on student achievement?
15. What effects, if any, will SOL testing have on students with disabilities?
16. What effects, if any, will SOL testing have on economically disadvantaged students?

Beliefs about the effects of SOL testing on *outcomes for instructional practices.*

2. To what degree, if any, will SOL testing affect the way that teachers teach?
4. To what degree, if any, will SOL testing affect the curriculum that teachers emphasize?
7. To what degree, if any, will SOL testing affect the amount of time that will be spent on test preparation?
10. To what degree, if any, will SOL testing affect the teaching of critical thinking?

Beliefs about the effects of SOL testing on *outcomes for teachers.*

5. To what degree, if any, will SOL testing affect teachers' job satisfaction?
11. To what degree, if any, will SOL testing affect the degree of stress that teachers feel in their jobs?
14. To what degree, if any, will SOL testing affect teachers' plans to remain in teaching?

The 11 questions were asked in numerical order: 1, 2, 4, 5, 7, 10, 11, 12, 14, 15, 16. Numbers were retained from the February 2001 content validity schedule in Appendix G.

Interview Protocol

"Hello, _____, this is Dale Margheim. I am a school principal and a doctoral student at Virginia Tech. You recently responded to my survey on the likely effects of high-stakes testing in Virginia, and you indicated that I could contact you for a telephone interview. Is this a good time for us to talk, or should I call you at a later time?"

"This part of my study will ask randomly-selected respondents to my survey for their opinions to some open-ended questions about the likely effects of Virginia's SOL testing program. The results of interviews will be included as part of my study."

"As required by Virginia Tech, be assured that you will not be identified in any way as a result of your participation in this study. Your responses will be attributed only to your respondent number."

"I will be taking notes to record your thoughts, but I can only write so fast. So, I do request your permission to tape record our interview. Tapes will be transcribed to ensure that all your thoughts are included. Tapes will then be destroyed."

"May I have permission to record our interview?"

"We are now beginning the interview with respondent number _____."

Volunteer educators classified each question into one of the three domains and rated each question for level of association and clarity. The educators made recommendations for rewording of questions for clarity or accuracy. These processes were accomplished in July of 2000 and again in October of 2000. Means and standard deviations were computed for the level of association of each question with the domain chosen and for the general clarity of each question. This process allowed me to improve the survey by reallocating questions to different domains or by rewriting, rewording, or deleting questions. The final content validity data for the interview portion of this research (February 2001) are in Appendix G.

Selection of Participants for Interviews

The selection of teachers to be interviewed was accomplished by assigning respondents agreeing to participate in phone interviews to a twelve cell sampling matrix which included school SES, SOL test grade status, and tenure status. School socio-economic status (SES) was derived from 1999-2000 free and reduced price lunch statistics (*www.pen.k12.va.us)*. I assigned teachers to three categories according to the estimated wealth of their school's community: (1) those with relatively low incidence of poverty (0-24% free and reduced price lunch eligibility), (2) those with moderate incidence of poverty (25-49% free and reduced price lunch eligibility), and (3) those with relatively high incidence of poverty (50% or more free and reduced price lunch eligibility). The selection matrix is in chapter 4. Representative numbers of teachers were randomly selected to be interviewed from each cell of the resulting selection matrix. This process helped to ensure that responses to interview questions represent the range of circumstances in which teachers teach and form their opinions. Characteristics of the interview population and sample are reported in Chapter 4.

Administration of Interviews

I tape-recorded the responses of all participants in the telephone interviews after gaining their permission to do so. Recorded interviews were transcribed to produce the exact words of interviewees. When placing each call, I identified myself and explained the purpose of the call and that the respondent had previously indicated an interest in being interviewed for the study. Each respondent was asked if that time would be convenient for an interview or if a second call at a later time would be more convenient. An interview time of 10 minutes was predicted based on trial interviews; however, the average time for each interview was approximately 30 minutes.

Lincoln and Guba (1985) explained the circumstances in which data-gathering may cease: (1) exhaustion of sources (no new sources to consult), (2) saturation of categories (continued interviewing produces no significant new categories), (3) emergence of regularities (interviews produce congruent responses), and (4) over extension, "…the sense that new information being unearthed is very far removed from the core of any of the viable categories that have emerged" (p. 350). I used these guidelines to help determine when the interview portion of the study should be concluded.

Analysis of Interview Data

Interview data were analyzed with the Constant Comparative Method (Maykut & Morehouse, 1994). Maykut and Morehouse stated that this method "combines inductive category coding with a simultaneous comparison of all units of meaning obtained" (p. 134). The Constant Comparative Method involves the following steps: (1) Typed transcriptions are photocopied, and the copies are marked to identify each unique unit of meaning. Units of meaning may be single short responses or even full paragraphs. Indeed, some units may even be longer such as when a respondent is telling a story. (2) Next each transcription is cut apart into individual units of meaning. Each unit is then taped to an index card, and each card is labeled to identify it with a specific respondent and the specific page of the transcript from which the thought was taken. (3) The next step, *discovery*, requires the researcher to carefully reread the transcriptions and to identify emerging themes in the data, perhaps expressed as phrases, questions, or recommendations. Category development proceeds continuously as new concepts appear in the transcriptions. As each new idea or concept is recorded, it is compared to existing categories. The idea may become part of an existing category, it may become part of a new category, it may become part of a miscellaneous category, or it may cause one or more categories to be combined.

The next step, *inductive category coding*, requires the researcher to review recurring concepts and themes to identify provisional categories into which virtually all thought units may eventually be grouped. The researcher continues the process with the *refinement of categories* by writing propositional statements that explain the criteria for inclusion in particular categories. Data cards are coded to ensure identification with a specific category. The next step, *exploration of relationships,* involves examining the various propositional statements to determine which will stand alone and which are meaningfully related to others. Maykut and Morehouse (1994) defined these as outcome propositions. The final step, *integration of data,* occurs when the researcher

combines the insights in various propositional statements and supporting units of thought into a summary with extensive use of supporting details emphasizing a rich description of respondents' views through quotes, vignettes, stories, and examples. This process is aided by use of raw data matrices in which data are summarized and categorized. The raw data matrices for this study are in Appendix H.